REMARKS/ARGUMENTS

1. Amendments

Claims 1, 4-6, 9, 10,11-14, 16-34, 40-49 are in the application. Claims 45-49 have been withdrawn.

Applicant has canceled claims 2, 3, 7, 8, 15, 35-39 and 50-59 without prejudice to the filing of a continuation application directed to the canceled subject matter.

Applicant has amended claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence comprising at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor Support for the recitation of at least 85% identity can be found in the specification and in original Claim 15. Support for the recitation of myelin basic protein can be found in the specification and in original Claim 8. Recitation of "Specific for MBP" can be found in the specification at, for example, page 12, lines 24-26 and page 18, lines 9 - 12. Support for the recitation of 90% sequence identity with the autoantigen sequence can be found in the specification and in original claim 9.

Applicant has amended claim 10 to recite the fusion molecule of claim 9 wherein said autoantigen sequence present in said fusion molecule comprises the amino acid sequence of SEQ ID NO:13. Support for this amendment can be found in the specification at, for example, page 621, line 20 to page 62, line 4 and in Example 2, SEQ ID NO:13.

Applicant has amended claims 4 - 6, 9,11-12, 16-17, 26 and 43 to clarify the claim language and correct claim dependencies. No new matter is added by these amendments.

2. Restriction Requirement

Applicant elected group 1 with traverse:

Group 1. Claims 2 - 34 and 40-44, drawn to an isolated fusion molecule wherein the autoantigen is **myelin basic protein**, pharmaceutical composition and kit comprising said fusion protein, classified in Class 530, subclass 350; Class 424, subclass 192.1, Class 435, subclass 810.

The Office Action indicated that claims 2-34 and 40-44 and claims 45-49 are related as product and process of use. Applicant thanks the office for the indication that where Applicant elects claims directed to a product and a product claim is subsequently found allowable, withdrawn process claims that depend from or otherwise include all of the limitations of the allowable product claim will be rejoined in accordance with the provisions of MPEP 821.4.

3. Objection to Disclosure

The disclosure stands objected to on the basis that there is a typographical error "pAN1782" on page 78, line 22. Applicant has amended the specification to correct this error and withdrawal of this objection is respectfully requested.

4. Rejection of Figure 9

Figure 9 stands rejected because the half-tone of the figure is allegedly too dark. Applicant hereby provides a new Figure 9. Withdrawal of this rejection is respectfully requested.

5. Rejection under 35 U.S.C. § 112, first paragraph. (Enablement)

Claims 1-34 and 40-44 stand rejected under 35 U.S.C. § 112, first paragraph because the specification, while being enabling for (1) isolated fusion molecule comprising hinge CH2-CH3 of human IgG1 constant region consisting of SEQ ID NO:1

fused to a full length myelin basic protein comprising SEQ ID NO:12 or a peptide from myelin basic protein consisting of SEQ ID NO:13, and (2) an isolated fusion molecule comprising hinge-CH2-CH3 of human IgG1 constant region consisting of SEQ ID NO:1 fused to human IgE constant region CH2-CH3-CH4 domains for inhibiting IgE mediated release of histamine, does not reasonably provide enablement for any fusion molecule as set forth in claims 1-34 and 40-44.

Applicant has canceled claims 2, 3, 7, 8, 15, 35-39 and 50-59 without prejudice to the filing of a continuation application directed to the canceled subject matter.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence comprising at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor

Applicant traverses the rejection for the following reasons.

The test for enablement entails an analysis of whether one skilled in the art is able to practice the invention using information disclosed in the application and information known in the art without undue or unreasonable experimentation (MPEP § 2164.01; see In *re Wands,* 858 F.2d 731, 8 USPQ 2d 1400, [Fed. Cir. 1988]). A finding of lack of enablement and determination that undue experimentation is necessary requires an analysis of a variety of factors *(i.e.,* the *In re* Wands factors). The most important factors that must be considered in this case include 1) the nature of the invention; 2) the level of ordinary skill in the art; 3) guidance provided in the specification, and 4) the state of the prior art. "[H]ow a teaching is set forth, by specific example or broad terminology, is not important"; and furthermore still, "[l]imitations and examples in the specification do not generally limit what is covered by the claims" (MPEP § 2164.08). The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art. *Ansul Co. v. Uniroyal, Inc.* 448 F.2d 872, 878-79; 169 USPQ 759, 762-63 (2d Cir. 197 1), cert. denied, 404 U.S. 10 18,

30 L. Ed. 2d 666, 92 S. Ct. 680 (1972). The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed. It is well settled that patent applicants are not required to disclose every species encompassed by their claims, even in an unpredictable art. The legal standard merely requires that there must be sufficient disclosure, either through illustrative examples or terminology, to teach those of ordinary skill how to make and use the invention as broadly as it is claimed. *Enzo Biochem., Inc. v. Calgene, Inc.,* 188 F.3d 13 62 (Fed. Circ. 1999), at 1372 (quoting *In re Vaeck*, 947 F.2d 488, 496 (Fed. Cir. 1991).

Proper application of the legal standard must lead to the conclusion that all claims pending in this application are fully enabled.

The nature of the invention

The present invention concerns certain novel fusion molecules that are capable of cross-linking a native IgG inhibitory receptor with a native IgE receptor. The fusion molecules comprise a sequence comprising 85% identity to IgG heavy chain sequence linked to a polypeptide autoantigen sequence which comprises 90% identity to myelin basic protein and is capable of being specifically bound by an immunoglobin specific for myelin basic protein. The purpose of these molecules is to allow the myelin basic peptide to function as an immunogen while any fusion peptides that reacted with IgE loaded mast cells would not trigger an adverse reaction. Such molecules find utility in the management of multiple sclerosis. While the therapeutic strategy underlying the present invention is both novel and unobvious, the fusion molecules themselves have a relatively simple structure, and can be made and tested by standard techniques that were well known in the art at the time of making the present invention. Furthermore, at the time the present invention was made, there was a lot of information known in the art about the interaction of IgG inhibitory receptors and IgE receptors with antibody constant regions, which provides valuable information for the construction of the fusion molecules of the present invention. Accordingly, although unpredictability in the field of

recombinant DNA technology is generally viewed as relatively high, the unpredictability in the particular field to which the present invention pertains is of lesser degree.

The level of ordinary skill in the art

It is well established that the level of skill in the art of recombinant DNA technology is relatively high, and is typically represented by the knowledge of a Ph.D. scientist with several years of experience in the pertinent field.

Fusion Molecules Comprising IgG heavy chain constant regions and peptide sequences with identity to myelin basic protein

The Examiner reasons that fusion molecules "comprising" first and second polypeptide sequences are not enabled, as use of the term "comprising" expands the hinge of a human IgG1 constant region to include additional amino acids at either ends of the first polypeptide within the claimed fusion molecule. There is allegedly a lack of guidance as to which amino acids are to be included. Further the first polypeptide would include the whole IgG rather than just the constant region of the human IgG1 because the term "comprise" is recited in said claims. (Office Action, page 8, paragraph starting on line 1).

Applicant must respectfully disagree. The present application describes, by way of example, additional non-essential but advantageous amino acid sequences and other elements that find use with the first and second polypeptides of the fusion molecules of the invention. For example, the first and second polypeptide sequences of the fusion molecule can be joined using various linkers (described in the Specification at page 56, lines 4-16). Also, the fusion molecules may contain posttranslational modifications, either naturally occurring or artificial, for example, acetylation, glycosylation and prenylation (see Specification page 21, lines 4 - 24). The Specification teaches that fusion polypeptide variants can be constructed that contain advantageous insertions of various amino acid sequences (page 21, line 25 to page 23, line 3), resulting in fusion molecules that have improved affinity for their respective IgG or IgE Fc receptors (Specification, page 34, line 24 to page 35, line 25). The fusion molecules of the invention can also comprise multiple copies of the IgG and

autoantigens, as described in page 54, lines 18-21. Fusion polypeptides further comprising signal sequences for intracellular localization or extracellular export (page 63, lines 20-22), and peptide sequence tags to facilitate fusion molecule purification (page 63, line 32 to page 64, line 3) also find use with the fusion molecules of the invention.

As outlined above, the Specification provides sufficient guidance to make a variety of advantageous fusion molecules comprising first and second polypeptide sequences. Applicant submits that fusion molecules comprising first and second polypeptides are fully enabled in view of 1) guidance provided throughout the Specification[1] (as described above), 2) the routine nature of recombinant DNA engineering and the production of chimeric or variant polypeptides, as known in the art, and 3) the high level of technical competence of one of ordinary skill in the immunological, genetics and protein-chemistry arts. The routine nature of manipulation of DNA and protein molecules is well known, as evidenced by the publications cited in the Specification *(see,* especially, page 20, line 29 to page 21, line 24; page 64, lines 17 - 26). Detailed protocols for the construction of the fusion molecule variants described in the Specification is not necessary for one of ordinary skill to practice the claimed invention without undue experimentation.

The Examiner asserts that use of the term "comprising" in the claims results in an infinite number of possible sequences, and it is not possible to predict which of those molecules will have desired functional activity. The Examiner asserts that there is also insufficient guidance as to which "portion" of which autoantigen in the fusion molecule is effective for treating which autoimmune disease. Applicant disagrees. Applicant points out that Claim 1 and all claims dependent on Claim 1 contain the functional limitation that the IgG domain has the ability to bind to the native IgG inhibitory receptor and that the myelin basic protein (MBP) is capable of specifically binding by a third polypeptide to a native IgE receptor.. The Specification teaches which amino acids are necessary for IgG receptor binding (see page 35, lines 1 - 25) as well as methods to determine the

[1] Applicant points out that the guidance provided in the Specification is found both in the Experimental Example as well as in the description of other preferred embodiments elsewhere in the Specification.

affinity of an Fc domain for its cognate receptor (see, for example, page 55, lines 15-25). Thus, use of the term "comprising" does not result in an infinite number of fusion molecules with unpredictable activities, and the identification of constructs that meet the limitation of the claims does not require undue experimentation.

The specification defines the term "portion" on page 29, lines 21 - 30. The specification teaches which amino acids comprise the epitopes of MBP (page 29, lines 7 to 11 and page 46, Table 2).

Applicant submits that use of the open-ended transitional phrase "comprising" is appropriate, that the term "portion" is appropriate and that the claims are enabled and commensurate in scope with the disclosure, and are allowable. The Examiner is respectfully requested to withdraw this rejection.

Fusion Molecules Comprising First Polypeptide Sequences Having 85% identity with an Ig Heavy Chain Constant Region and comprising Second Polypeptide Sequences having 90% identity to myelin basic protein are Enabled

The Examiner alleges that the disclosure of the Specification provides insufficient guidance to make and use polypeptides having at least 85% sequence identity with the IgG constant domain sequences of SEQ ID NO: 3, where the molecules retain biological or immunological function. The Examiner cites Attwood et al., and Skolnick et al., for the premise that knowing a protein's structure does not tell one its function.

Applicant respectfully traverses the rejection. Applicant points out that the Specification describes methods for the determination of percent identity between two amino acid sequences (see Specification, page 23, lines 4 - 13). In fact, the specification teaches specific parameters to be associated with the term "percent identity" as applied to the present invention. Also, one of ordinary skill in the art will recognize that the prior art provides numerous sources that describe IgG Fc sequences highly homologous to the Fc sequences of SEQ ID NO: 3 *(see,* the Specification at page 26, line 19 - page 27, line 6). Furthermore, Applicant asserts that one of ordinary skill in the art has a sufficiently high level of technical competence to experimentally identify novel Fc sequences having at least 85% sequence identity with the constant

domain sequences of SEQ ID NO: 3 using hybridization methods provided in the Specification (see, page 23, line 16 to page 24, line 12). Alternatively, one of skill in the art can readily engineer novel Fc domains having at least 85% sequence identity with the constant domain sequences of SEQ ID NO: 3 using recombinant DNA/protein engineering techniques. Thus, detailed protocols for the construction of fusion molecules having at least 85% sequence identity with a known Ig constant domain in the Specification is not necessary in order for one of ordinary skill to practice the claimed invention without undue experimentation.

Applicant points out that all pending claims reciting polypeptides having at least 85% sequence identity with IgG domains *(e.g.,* having 85% sequence identity with SEQ ID NO: 3) contain the functional limitation that the polypeptides also have the ability to bind to their respective cell surface receptors, and the Specification teaches which amino acids are necessary for receptor binding (see page 35, lines 1-25) as well as methods to determine the affinity of an Fc domain for its cognate receptor (see, for example, page 55, lines 15 - 25). Conservation of these critical amino acids results in polypeptides that retain the desired biological activity (receptor binding and inhibition of histamine release). Thus, contrary to the Examiner's statement, use of the phrase "having at least 85% sequence identity" does not result in fusion molecules with unpredictable activities. A person skilled in the art, relying on the teaching of the specification and general knowledge in the art at the time the present invention was made, will be able to determine which amino acid alterations are expected to yield constructs that satisfy the functional limitations of the claims, without undue experimentation.

Moreover, it is well-known in the art that many, if not most, polypeptides of the invention having at least 85% sequence identity with the constant domain sequence of SEQ ID NO: 3 will retain biological activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid). For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in J. U. Bowie, *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," Science 247:1306-10

(1990) (copy enclosed), wherein the authors indicate that proteins are surprisingly tolerant of amino acid substitutions.

In addition, amino acids in the fusion proteins of the present invention that are essential for function can be easily identified by methods well-known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (*see, e.g.*, Cunningham and Wells, *Science* 244:1081-85 (1989) (copy enclosed)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity as taught in the specification.

The claims currently recite peptide sequences associated with biological activity. This biological activity with the well defined relatively high degree of sequence identity and general knowledge in the art at the time the invention was made, is believed to sufficiently define the claimed genus such that, one skilled in the art, at the effective date of the present application, would have known how to make and use the claimed peptide sequences without undue experimentation. As the M.P.E.P. states, "[t]he fact that experimentation may be complex does not necessarily make it undue, if the art typically engages in such experimentation," *In re Certain Limited-charge Cell Culture Microcarriers,* 221 USPQ 1165, 1174 (Int'l Trade Comm'n 1983), *aff' sub nom. Massachusetts Institute of Technology v A.B. Fortia* 774 F 2d 1104, 227 USPQ 428 (Fed. Cir. 1985);*M.P.E.P. 2164.01*

Lack of in vivo working examples

The Examiner indicates that there is a lack of *in vivo* working examples demonstrating that the fusion molecule is effective for treating multiple sclerosis. The Examiner relies on Blanas et al., Couzin et al. and Mackay et al[2]. The Examiner states that the fusion molecule may be inactivated before producing an effect; the fusion molecule may not reach its targeted area.

The legal standard with respect to *in vitro* or animal model data providing pharmacological activity has been commented on in *Cross v Iizuka* 753 F. 2d 1040, 1051, 224 USPQ 739, 747-48 (Fed. Cir 1985).

---

[2] Applicant presumes that the Examiner means Davidson et al., (2001) New England J. of Med. 345(5) 340-350, Eds. MacKay & Rosen rather than MacKay et al.

"We perceive no insurmountable difficulty, under appropriate circumstances, in finding that the first link in the screening chain, *in vitro* testing, may establish a practical utility for the compound in question. Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vitro* utility."

Furthermore, M.P.E.P. 2107.03 (III) states

"If reasonably correlated to the particular therapeutic or pharmacological utility, data generated using *in vitro* assays, or from testing in an animal model or a combination thereof almost invariably will be sufficient to establish therapeutic or pharmacological utility for a compound, composition or process."

Thus, the legal standard accepts that *in vitro* or animal model data is acceptable utility as long as the data is "reasonably correlated" to the pharmacological utility described.

Blanas indicates that oral administration of ovalbumin autoantigen in mice was found to induce a cytotoxic T lymphocyte response that could lead to the onset of autoimmune diabetes. Blanas does not discuss the MBP peptide or multiple sclerosis. Applicant's construct comprises the heavy chain constant region of the IgG fused to the MBP peptide. The fusion molecule acts to inhibit the autoallergic reaction. Blanas does not discuss the administration of a fusion molecule, let alone a fusion molecule comprising an autoantigen fused to the IgG heavy chain constant region, as claimed. Accordingly, the findings of Blanas cannot be applied to the currently claimed invention.

Couzin et al. (2003) is an article reviewing various clinical tests for the treatment and prevention of type I diabetes. Couzin does not discuss the MBP peptide, use of a fusion polypeptide or multiple sclerosis. For the reasons set forth for Blanas, the findings of Couzin et al. cannot be applied to the currently claimed invention. Furthermore, for the reasons set forth above, the legal standard sufficient to establish utility of a compound is in vitro or in vivo tests. The legal standard is not human clinical trials.

Mackay et al., states that two recent phase I clinical trials for treatment of multiple sclerosis by administering altered peptide ligands derived from MBP resulted in

either hypersensitivity reactions or exacerbations of multiple sclerosis. (page 346) First, MacKay does not indicate that the altered peptide ligands derived from MBP are functionally attached to the IgG heavy chain constant regions. The purpose of the IgG Fc regions is to prevent the hypersensitivity reaction seen with the peptides as taught by MacKay. Accordingly, MacKay does not teach that the claimed invention will not work. Furthermore, for the reasons set forth above, the legal standard sufficient to establish utility of a compound is in vitro or in vivo tests. The legal standard is not human clinical trials.

Applicant encloses later published papers which show that fusion molecules comprising an IgG constant region linked to an IgE constant region successfully reduces histamine release in animals. Clearly such compounds are not inactivated as suggested by the Examiner. Clearly these types of fusion molecules can be successfully administered to animals[3].

Furthermore, the appearance of IgG or other antibodies against the MBP portion of the fusion molecule would not be a problem because the purpose of the molecule is to present the MBP as an "immunogen" while any reacted IgE loaded mast cells would be suppressed by the IgG Fc portion.

For the above reasons, the Examiner is respectfully requested to withdraw this rejection.

6. Rejection under 35 U.S.C. § 112, first paragraph. (Written Description)

Written Description

Claims 1-34 and 40-44 stand rejected under 35 U.S.C. § 112, first paragraph for allegedly lacking written description. Specifically, the Examiner alleges that there is insufficient written description in the Specification for the same fusion molecules that were rejected on the basis of lack of enablement (see above, and Office Action, pages 7-8).

---

[3] Zhu et al., "A novel human immunoglobulin Fcγ-Fcε bifunctional fusion protein inhibits Fcε RI mediated degranulation". (2002) *Nature Medicine* vol. 8 (5) 518-521

Applicant has canceled claims 2 - 3, 7 - 8, 15, 35 - 39 and 50 - 59 without prejudice.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence comprising at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

Applicant traverses the rejection for the following reasons.

To satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention (e.g., *Vas-Cath, Inc. v. Mahurkar,* 935 F.2d at 563, 19 USPQ 2d at 1116 and *Gentry Gallery, Inc. v. Berkline Corp.,* 134 F.3d 1473, 45 USPQ 2d 1498 [Fed. Cir. 1998]). Applicant asserts that they have met this requirement. Applicant emphasizes that sufficient written description must be ascertained in view of one skilled in the art. "It is not required that the application describe the claim limitations in greater detail than the invention warrants. The description must be sufficiently clear that persons of skill in the art will recognize that the applicant made the invention having those limitations" *(Martin v. Mayer,* 823 F.2d 500, 3 USPQ 2d 1333 [Fed. Cir. 1987]),

Multiple Fusion Molecules are Described in the Specification

The Examiner alleges that the Specification discloses insufficient written description of the structure of fusion molecules of the claimed invention to support a claim to a larger genus of fusion molecules.

Applicant must respectfully disagree. As described above, the Specification describes multiple fusion molecules. For example, the Specification describes the construction of chimeric fusion molecules, see Example 2, pages 180-183. The

---

Kepley et al. "FcεRI-FcγRII coaggregation inhibits IL-16 production from human langerhans-like dendritic cells" (2003) *clinical Immunology* vo. 108 p. 89-94

Specification also describes fusion molecules where the first and second polypeptide sequences of the fusion molecule are connected by use of linkers (see Specification page 27, lines 4-15). Also, the fusion molecules may contain post translational modifications, either naturally occurring or artificial, for example, acetylation, glycosylation or prenylation (as described in the Specification at page 21, line 4 - 24). The Specification describes advantageous fusion molecule variants (page 21, line line 25 - page 23, line 3), where the variants have improved affinity for their respective IgG or IgE receptors (Specification, page 34, line 24 - page 35, line 25). The Specification describes fusion molecules comprising multiple copies of IgG and autoantigen (page 54, lines 18-21). Fusion polypeptides further comprising signal sequences for intracellular localization or extracellular export (page 63, lines 20-22), and peptide sequence tags to facilitate fusion molecule purification the fusion molecules (page 63, line 32 to page 64, line 3) are also described.

In view of the fusion molecules described above and the level of skill in the art, Applicant asserts that sufficient representative fusion molecules are adequately described in the Specification (without undue detail) to support a genus of fusion molecules, as recited in Claim 1, and all claims dependent on Claim 1. Applicant respectfully requests the withdrawal of this rejection.

Fusion Molecules Comprising First and/or Second Polypeptide Sequences Having "85% Sequence Identity " with an Ig Heavy Chain Constant Region are Adequately Described

The Examiner alleges that the Specification fails to provide sufficient written description of polypeptides having at least 85% sequence identity with the IgG constant domain sequences *(e.g.,* 85% sequence identity with SEQ ID NO: 3) where the molecules retain biological activity.

Applicant respectfully traverses the rejection. Applicant points out that the Specification describes methods for the determination of percent identity between two amino acid sequences (see Specification, page 23, lines 4 - 13). Also, the Specification provides examples of prior art that describes numerous Ig Fc polypeptides having at least 85% sequence identity with the Fc sequences of SEQ ID NO: 3 *(see,* the

Specification at page 26, line 19 to page 27, line 6). The Specification also describes methods for the identification of Ig Fc sequences having at least 85% sequence identity with the constant domain sequences of SEQ ID NO: 3 (see the Specification at page 23, line 16 to page 24, line 12). Alternatively still, one of ordinary skill in the art can readily engineer novel Fc domains having at least 85% sequence identity with the constant domain sequences of SEQ ID NO: 3 using recombinant DNA/protein engineering techniques.

Applicant points out that all pending claims reciting polypeptides having at least 85% sequence identity with IgG Fc domains *(e.g.,* having 85% sequence identity with SEQ ID NOs: 3) contain the functional limitation that the polypeptides also have the ability to bind to the IgG cell surface receptor. The Specification provides description of this limitation where the amino acids necessary for receptor binding and biological activity (page 35, lines 1-25) and methods to determine the affinity of an Fc domain for its cognate receptor (see, for example, page 55, lines 15 - 25) are described.

Applicant argues that the Specification provides adequate written description for fusion molecules comprising polypeptides having at least 85% sequence identity with IgG constant domain sequences *(e.g.,* 85% sequence identity with SEQ ID NO: 3), especially in view of the state of the prior art, and the high level of skill in the art. Applicant further argues that the scope of the claims finds written description throughout the Specification, and are allowable. The Examiner is respectfully requested to withdraw this rejection.

7. Rejection under 35 U.S.C. § 112, second paragraph

Claims 12 and 13 stand rejected under 35 U.S.C.§ 112, second paragraph as being indefinite for failing to point out and distinctly claim the subject matter which applicant regards as the invention.

The "said IgE receptor" in claims 12 and 13 allegedly does not have antecedent basis in claim 11. The "FceR1 IgG receptor" in claim 12 is indefinite because the FceRI receptor is an IgE receptor.

Applicant has amended claims 11 and 12 to clarify the language of the claims. These amendments remove any basis for these rejections and withdrawal of the rejections is respectfully requested.

8. Rejection under 35 U.S.C. § 102(b)

Claims 1-4, 9-11, 15-16, 22-28, 32-34 and 40-41 stand rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 5,420,247 (May 30, 1995).

The office action states that the '247 patent allegedly teaches an isolated fusion molecule comprising a first polypeptide such as human IgG Fc constant region directly functionally connected to a second autoantigen polypeptide sequence such as human LIFR that is capable of specific binding to LIF ligand through a third polypeptide sequence, such as a polypeptide linker.

Applicant has cancelled claims 2-3 and 15, without prejudice. Withdrawal of the rejection with respect to these claims is requested.

Applicant traverses this rejection for the following reasons:

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of myelin basic protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor

The standard for anticipation under 35 U.S.C. § 102(b) is strict identity. Anticipation under § 102(b) can only be established by a single prior art reference that teaches each and every element of the claimed invention (*Structural Rubber Products Co. v. Park Rubber Co.* 223 USPQ 1264 (1984)).

The '247 patent does not teach a fusion molecule comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP). The'247 patent does not teach each and every element of the claimed

invention. Accordingly, the '247 patent does not anticipate claim 1 as amended. Withdrawal of this rejection is respectfully requested.

9. Rejection under 35 U.S.C. § 103(a) over U.S. Patent 5,420, 247 in view of WO95/14779, Basu et al. and Daeron et al.

Claims 1, 3 and 11-14 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,420,247 in view of WO95/14779, Basu et al., (J. Biol. Chem 268(18) 13118-27) and Daeron et al., (J. Clin Invest)

The Office Action states that it would have been obvious to one having ordinary skill in the art at the time the invention was made to substitute the LIF-R in the fusion molecule as taught by the '247 patent for the human IgE heavy chain constant region that binds to an high affinity IgE receptor as taught by the WO 95/14779 publication and Basu et al. and low affinity IgE receptor as taught by Daeron et al. for a fusion molecule comprising a first polypeptide IgG heavy chain constant region that is capable of binding to IgG inhibitory receptor such as low affinity FcγRIIb connected to an IgE heavy chain constant region sequence that is capable of binding to an IgE receptor such as high affinity receptor or low affinity receptor through a third polypeptide linker sequence to inhibit IgE induced mediator release.

Applicant traverses this rejection for the following reasons:

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of myelin basic protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

The '247 patent teaches LIF-R proteins. The '247 patent teaches that the LIF-R peptide may be fused to an antibody Fc region, which comprises the hinge region in order to allow dimerization of the peptide chains through binding of the cysteine residues in the hinge region.

WO95/14779 teaches mutated human IgE fragments comprising the second, third and fourth constant region domains of the IgE heavy chain. The publication teaches mutated glycosylated polypeptides which include at least a part of the IgE heavy chain sufficient to bind to the FcεRI or FcεRII receptor sites on human cells which are useful in the investigation and amelioration of allergic conditions. WO95/14779 does not teach or suggest directly linking the IgG heavy chain constant region with a myelin basic protein fragment in one molecule.

Basu *et al.*, teach that the Fc region of IgE comprising epsilon 2, epsilon 3 and epsilon 4 domains are sufficient for binding to the IgE high affinity receptor. Basu does not teach or suggest directly linking the Fc region of IgG with a myelin basic protein fragment in one molecule

Daeron *et al.* teach that linking the FcεRI receptor with the FcγRII low affinity receptor will result in inhibition of IgE induced release of mediator and cytokines. Daeron *et al.* accomplishes this linkage by establishing an artificial system, using a rat basophilic leukemia cell transfected with mouse FcγRII and naturally expressing rat FcεRI receptor. The cells were then exposed to three different antibodies to link the mouse FcγRII with the rat FcεRI receptor. These antibodies were rat anti-mouse FcγRII/III Fab'2; mouse anti-rat Ig Fab'2 and rat IgE. There is no teaching or suggestion in Daeron *et al.* to make a single molecule linking the Fc regions of IgG with a myelin basic peptide fragment.

The claimed invention is not obvious in light of the combination of the cited references for the following reasons.

Where claimed subject matter has been rejected as obvious in view of a combination of prior art references, a proper analysis under 35 U.S.C. §103 requires, inter alia, consideration of two factors: (1) whether the prior art would have suggested to those of ordinary skill in the art that they should make the claimed composition or device, or carry out the claimed process and (2) whether the prior art would also have revealed that in so making or carrying out, those of ordinary skill would have a reasonable expectation of success. *In re Vaeck* 947 F 2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

First, there is no teaching or suggestion in the combination of the references to fuse the heavy constant region of the IgG molecule with a myelin basic peptide fragment. While the '247 patent teaches LIF-R fused with IgG, the purpose of adding IgG was to generate a dimer molecule. Both WO95/14779 and Basu teach the region of the heavy chain of IgE which binds to the IgE receptor. Neither teaches or suggests fusing the heavy chain region of IgG with a myelin basic peptide fragment. Finally Daeron, while suggesting indirectly linking the FcεRI receptor with the FcγRII receptor, does not teach or suggest fusing the heavy chain region of IgG with a myelin basic peptide fragment.

Secondly, there is no motivation in the cited references to combine the teachings of the references to arrive at the claimed invention. The '247 patent uses the IgG heavy chain region to generate a dimer molecule to obtain the biologic function of LIF. It provides no motivation to replace the LIF-R4 peptide with an MBP peptide in which the MBP serves as an immunogen/tolergen. Daeron et al. discusses the advantages of linking the FcεRI receptor with the FcγRII receptor. However, Daeron proposes a solution to this using multiple antibody molecules. There is no motivation in Daeron to fuse the heavy chain of IgG with a myelin basic peptide. Accordingly, Daeron does not provide motivation to generate a fusion molecule comprising the heavy chain constant region of IgG fused to a myelin basic peptide. If anything, Daeron *et al.*, teaches away from the claimed invention by teaching that linking of the FcεRI receptor to the FcγRII receptor can be achieved through the use of multiple antibodies.

Finally, none of the references, either alone or in combination provides a reasonable expectation of success from the claimed invention.

Absent a suggestion in the art to make the claimed invention, a motivation in the cited references to combine the references into the claimed invention and a reasonable expectation of success, the claimed invention is not obvious. Withdrawal of this rejection is respectfully requested.

10. Rejection under 35 U.S.C. § 103(a) over U.S. Patent 5,420, 247 in view of Warren et al.

Claims 1, 3 and 7-8 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,420,247 in view of Warren et al. (Proc. Natl. Acad Sci USA 92: 11061-65 (1995)). It allegedly would have been obvious to substitute the LIF-R in the fusion molecule as taught by the '247 patent for the autoantigen such as myelin basic protein (MBP) as taught by Warren et al. for a fusion molecule comprising a first polypeptide IgG heavy chain constant region connected to myelin basic protein through a polypeptide linker as taught by the '247 patent and Warren et al. From the combined teachings, a skilled practitioner would allegedly have had a reasonable expectation of success in producing the claimed invention.

Applicant has canceled claims 3 and 7-8 without prejudice. Withdrawal of this rejection is requested for these claims.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor

The teachings of the '247 patent have been discussed. The '247 patent does not teach or suggest a fusion molecule of the IgG heavy chain constant region fused to an MBP peptide.

Warren et al., teach autoantigen sequences such as myelin basic protein and various antigenic epitopes or fragments thereof such as MBP84-93. Warren et al. teach that free and/or cerebrospinal fluid tissue bound autoantibodies to MBP are found in a large number of patients with MS and optic neuritis. Increased anti-MBP levels are highly associated with disease activity. Warren does not teach or suggest fusion molecules of the MBP peptide with the IgG heavy chain constant region. Warren does not teach or suggest the administration of an IgG Fc-MBP peptide fusion polypeptide to persons for the treatment of multiple sclerosis.

The claimed invention is not obvious for the following reasons.

Where claimed subject matter has been rejected as obvious in view of a combination of prior art references, a proper analysis under 103 requires, *inter alia,* consideration of two factors: (1) whether the prior art would have suggested to those of ordinary skill in the art that they should make the claimed composition or device, or carry out the claimed process and (2) whether the prior art would also have revealed that in so making or carrying out, those of ordinary skill would have a reasonable expectation of success. *In re Vaeck* 947 F 2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

First, there is no teaching or suggestion in the combination of the references to fuse the heavy constant region of the IgG molecule with a myelin basic peptide fragment. While the '247 patent teaches LIF-R fused with IgG, the purpose of adding IgG was to generate a dimer molecule. While the Warren et al. reference teaches MBP peptides there is no teaching or suggestion to combine the peptide with an IgG Fc region.

Secondly, there is no motivation in the cited references to combine the teachings of the references to arrive at the claimed invention. The '247 patent uses the IgG heavy chain region to generate a dimer molecule to obtain the biologic function of the LIF molecule. It provides no motivation to replace the LIF-R4 peptide with an MBP peptide in which the MBP serves as an immunogen/tolergen. There is no motivation in Warren et al. to attach the MBP peptide to an IgG heavy region. There is no motivation to generate a compound comprising the MBP peptide.

Finally, none of the references, either alone or in combination provide a reasonable expectation of success from the claimed invention.

Absent a suggestion in the art to make the claimed invention, a motivation in the cited references to combine the references into the claimed invention and a reasonable expectation of success, the claimed invention is not obvious. Withdrawal of this rejection is respectfully requested.

11. Rejection under 35 U.S.C. § 103(a) over U.S. Patent 5,420,247 in view of U.S. Patent 5,565,335

Claims 1-4, 15 and 18-21 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,420,247 in view of U.S. Patent 5,565,335. The Office Action states that it would have been obvious to one having ordinary skill in the art at the time to substitute the Fc polypeptide in the fusion protein as taught by the '247 patent for the human IgG1 Fc having an amino acid sequence at lest 97.2% identical to the claimed SEQ ID NO:3 for a fusion molecule comprising a human IgG Fc constant region functionally connected to any second autoantigen polypeptide sequence such as a human LIFR, or myelin associated protein or CD4 through a third polypeptide sequence such as a polypeptide linker. One having ordinary skill in the art would allegedly be motivated to do this because the '247 patent teaches that IgG Fc fusion molecule can be easily purified and the '335 patent teaches that Fc fusion molecules enhance the plasma half-life of the fusion molecule.

Applicant has canceled claims 2-3 and 15 without prejudice. Withdrawal of this rejection is requested for these claims.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

U.S. Patent 5,420,247 has already been discussed.

U.S. Patent No. 5,565,335 teaches soluble secreted adhesons comprising the CD4 protein. The CD4 adheson ordinarily binds to the recognition sites of HIV and the purpose of the patent is to design candidates for therapeutically sequestering these HIV sites, thereby blocking viral infectivity. The '335 patent teaches fusing the CD4 polypeptide with a protein with a long plasma life such as an immunoglobulin constant domain. The purpose of this fusion is to increase the half-life of the CD4 polypeptide. The '335 patent teaches the CD4 peptide linked to the IgG1 heavy chain constant region. The '335 patent teaches that adhesons are cell surface polypeptide having an extra-cellular domain which is homologous to a member of the immunoglobulin gene

superfamily, excluding however, highly polymorphic members of the superfamily. (Col. 4, lines 7 - 14). The patent lists a number of examples of adhesions. Myelin basic protein is not an adheson. There is no teaching or suggestion of autoantigens. There is no teaching or suggestion in the '335 patent to replace the CD4 molecule in the immunoadheson with an MBP peptide. Such a replacement would be against the purpose of the '335 patent.

The claimed invention is not obvious in light of the references for the following reasons.

First, there is no teaching or suggestion in the combination of the references to fuse the heavy constant region of the IgE molecule with a myelin basic peptide fragment. While the '247 patent teaches LIF-R fused with IgG, the purpose of adding IgG was to generate a dimer molecule. While the '335 patent teaches CD4 fused with IgG there is no teaching or suggestion to combine an MBP peptide with an IgG Fc region.

Secondly, there is no motivation in the cited references to combine the teachings of the references to arrive at the claimed invention. The '247 patent uses the IgG heavy chain region to generate a dimer molecule to obtain the biologic function of the LIF molecule. It provides no motivation to replace the LIF-R4 peptide with an MBP peptide. The '335 patent teaches CD4 fused with IgG. There is no motivation in the '335 patent to replace the CD4 with an MBP peptide.

Finally, none of the references, either alone or in combination provide a reasonable expectation of success from the claimed invention.

Absent a suggestion in the art to make the claimed invention, a motivation in the cited references to combine the references into the claimed invention and a reasonable expectation of success, the claimed invention is not obvious. Withdrawal of this rejection is respectfully requested.

12. Rejection under 35 U.S.C. § 103(a) over U.S. Patent 5,420, 247 in view of Elias et al. and Marks et al.

Claims 1 and 29-31 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,420,247 in view of Elias et al., (J. Biol. Chem

265(26) 15511-17, (1990) and Marks et al., (J. Cell Biol. 135(2) 341-354, (1996). The Office action states that it would have been obvious to add at least one amino terminal ubiquitination target motif such as large hydrophobic amino acid residue such as leucine as taught by Elias and Marks to the fusion molecule as taught by the '247 patent to target the transmembrane protein such as IgFc connected to LIF-R through a peptide linker to route the fusion molecule to the lysosome and endosome antigen processing as well as modulating the half-life of the fusion molecule as taught by the '247 patent, Elias et al. and Marks et al.

This rejection is traversed for the following reasons.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

The claimed invention is not obvious in light of the combination of the cited references for the following reasons.

The '247 patent has already been discussed.

Elias et al. teach N terminal residue of the protein is one important structural determinant recognized by ubiquitin ligase to conjugated protein to ubiquitin for protein degradation. Elias et al. teach hydrophobic amino acid residues such as leucine or basin amino acid residues such as histidine, arginine and lysine determine the half-life of the protein.

Mark et al. teach that adding ubiquitination target motifs such as bulky hydrophobic group di-leucine motifs to any protein would target the protein to the lysosome or endosomal compartments for antigen processing.

Where claimed subject matter has been rejected as obvious in view of a combination of prior art references, a proper analysis under 103 requires, inter alia, consideration of two factors: (1) whether the prior art would have suggested to those of ordinary skill in the art that they should make the claimed composition or device, or

carry out the claimed process and (2) whether the prior art would also have revealed that in so making or carrying out, those of ordinary skill would have a reasonable expectation of success. *In re Vaeck* 947 F 2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

First, there is no teaching or suggestion in the combination of the references to fuse the heavy constant region of the IgG molecule with a myelin basic peptide fragment. The '247 patent does not teach or suggest the fusion of a heavy constant region of the IgG molecule with a myelin basic peptide fragment. The Elias and Marks references do not cure this deficiency.

Secondly, there is no motivation in the cited references to combine the teachings of the references to arrive at the claimed invention. The '247 patent uses the IgG heavy chain region to generate a dimer molecule. It provides no motivation to replace the LIF-R4 peptide with an MBP peptide. The Elias and Marks references do not cure this deficiency.

Finally, none of the references, either alone or in combination provide a reasonable expectation of success from the claimed invention.

Absent a suggestion in the art to make the claimed invention, a motivation in the cited references to combine the references into the claimed invention and a reasonable expectation of success, the claimed invention is not obvious. Withdrawal of this rejection is respectfully requested.

13. Rejection under 35 U.S.C. § 103(a) over U.S. Patent 5,420, 247 in view of U.S. Patent No. 5,945,294

Claims 1, 3 and 42-44 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,420,247 in view of U.S. Patent No. 5,945,294. The Office Action states that it allegedly would have been obvious to substitute the human Fc epsilon receptor as taught by the '294 patent for the fusion protein as taught by the '247 patent in the kit for diagnostic assays.

Applicant has canceled claim 3 without prejudice. Withdrawal of this rejection is requested for this claim.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain

constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

The '247 patent has been discussed.

The '294 patent teaches diagnostic kits for IgE detection comprising human Fc epsilon receptor and an allergen.

Neither the "247 nor the '294 patent nor a combination of both teaches or suggests the fusion protein of the IgG Fc region with the MBP peptide in a kit. Absent such a teaching or suggestion, the invention is not obvious.

Withdrawal of this rejection is respectfully requested.

14. Rejection under judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 77, 79-81, 83-89 and 96 of copending Application No. 09/847,208.

Claims 1-3, 11-28, and 40-44 stand provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 77, 79-81, 83-89 and 96 of copending Application No. 09/847,208.

Applicant has amended Claim 1 to recite an isolated fusion molecule comprising a first polypeptide sequence having at least 85% identity with an IgG heavy chain constant region sequence capable of specific binding to a native IgG inhibitory receptor, directly functionally connected to a second polypeptide autoantigen sequence comprising at least 90% sequence identity to at least a portion of the amino acid sequence of basic myelin protein (MBP) and capable of specific binding, through a third polypeptide sequence specific for myelin basic protein, to a native IgE receptor.

Applicant has canceled claim 3.

This rejection is traversed for the following reason. Application No. 09/847,208. does not claim a fusion molecule comprising a IgG Fc region fused to a myelin basic peptide fragment.

Withdrawal of this rejection is respectfully requested.

Applicant believes that this application is in condition for allowance.

Please direct any calls in connection with this application to the undersigned at the number provided below.

Respectfully submitted,

Date: March 22, 2005

Leslie A. Mooi
Reg. No. 37,047

HELLER EHRMAN WHITE & McAULIFFE LLP
Customer No. 25213
275 Middlefield Road
Menlo Park, California 94025
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

SV 2102443 v3
3/22/05 2:58 PM (39754.0674)

Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions

JAMES U. BOWIE,* JOHN F. REIDHAAR-OLSON, WENDELL A. LIM, ROBERT T. SAUER

An amino acid sequence encodes a message that determines the shape and function of a protein. This message is highly degenerate in that many different sequences can code for proteins with essentially the same structure and activity. Comparison of different sequences with similar messages can reveal key features of the code and improve understanding of how a protein folds and how it performs its function.

THE GENOME IS MANIFEST LARGELY IN THE SET OF PROteins that it encodes. It is the ability of these proteins to fold into unique three-dimensional structures that allows them to function and carry out the instructions of the genome. Thus, comprehending the rules that relate amino acid sequence to structure is fundamental to an understanding of biological processes. Because an amino acid sequence contains all of the information necessary to determine the structure of a protein (*1*), it should be possible to predict structure from sequence, and subsequently to infer detailed aspects of function from the structure. However, both problems are extremely complex, and it seems unlikely that either will be solved in an exact manner in the near future. It may be possible to obtain approximate solutions by using experimental data to simplify the problem. In this article, we describe how an analysis of allowed amino acid substitutions in proteins can be used to reduce the complexity of sequences and reveal important aspects of structure and function.

Methods for Studying Tolerance to Sequence Variation

There are two main approaches to studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. This method has been extremely powerful for proteins such as the globins or cytochromes, for which sequences from many different species are known (*2–7*). The second approach uses genetic methods to introduce amino acid changes at specific positions in a cloned gene and uses selections or screens to identify functional sequences. This approach has been used to great advantage for proteins that can be expressed in bacteria or yeast, where the appropriate genetic manipulations are possible (*3, 8–11*). The end results of both methods are lists of active sequences that can be compared and analyzed to identify sequence features that are essential for folding or function. If a particular property of a side chain, such as charge or size, is important at a given position, only side chains that have the required property will be allowed. Conversely, if the chemical identity of the side chain is unimportant, then many different substitutions will be permitted.

Studies in which these methods were used have revealed that proteins are surprisingly tolerant of amino acid substitutions (*2–4, 11*). For example, in studying the effects of approximately 1500 single amino acid substitutions at 142 positions in *lac* repressor, Miller and co-workers found that about one-half of all substitutions were phenotypically silent (*11*). At some positions, many different, nonconservative substitutions were allowed. Such residue positions play little or no role in structure and function. At other positions, no substitutions or only conservative substitutions were allowed. These residues are the most important for *lac* repressor activity.

What roles do invariant and conserved side chains play in proteins? Residues that are directly involved in protein functions such as binding or catalysis will certainly be among the most conserved. For example, replacing the Asp in the catalytic triad of trypsin with Asn results in a 10^4-fold reduction in activity (*12*). A similar loss of activity occurs in λ repressor when a DNA binding residue is changed from Asn to Asp (*13*). To carry out their function, however, these catalytic residues and binding residues must be precisely oriented in three dimensions. Consequently, mutations in residues that are required for structure formation or stability can also have dramatic effects on activity (*10, 14–16*). Hence, many of the residues that are conserved in sets of related sequences play structural roles.

Substitutions at Surface and Buried Positions

In their initial comparisons of the globin sequences, Perutz and co-workers found that most buried residues require nonpolar side chains, whereas few features of surface side chains are generally conserved (*6*). Similar results have been seen for a number of protein families (*2, 4, 5, 7, 17, 18*). An example of the sequence tolerance at surface versus buried sites can be seen in Fig. 1, which shows the allowed substitutions in λ repressor at residue positions that are near the dimer interface but distant from the DNA binding surface of the protein (*9*). These substitutions were identified by a functional

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

*Present address: Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90024.

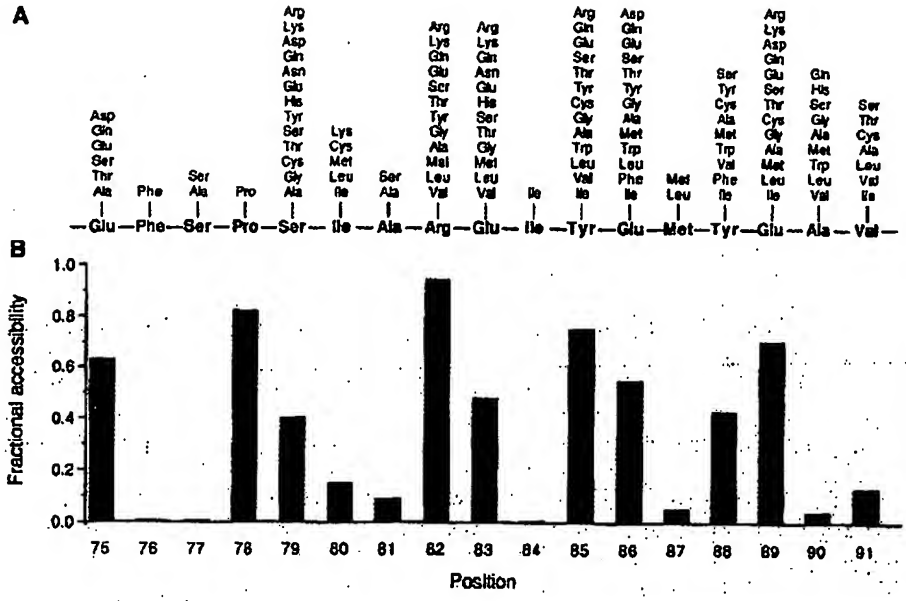

Fig. 1. (**A**) Amino acid substitutions allowed in a short region of λ repressor. The wild-type sequence is shown along the center line. The allowed substitutions shown above each position were identified by randomly mutating one to three codons at a time by using a cassette method and applying a functional selection (*9*). (**B**) The fractional solvent accessibility (*42*) of the wild-type side chain in the protein dimer (*43*) relative to the same atoms in an Ala-X-Ala model tripeptide.

selection after cassette mutagenesis. A histogram of side chain solvent accessibility in the crystal structure of the dimer is also shown in Fig. 1. At six positions, only the wild-type residue or relatively conservative substitutions are allowed. Five of these positions are buried in the protein. In contrast, most of the highly exposed positions tolerate a wide range of chemically different side chains, including hydrophilic and hydrophobic residues. Hence, it seems that most of the structural information in this region of the protein is carried by the residues that are solvent inaccessible.

Constraints on Core Sequences

Because core residue positions appear to be extremely important for protein folding or stability, we must understand the factors that dictate whether a given core sequence will be acceptable. In general, only hydrophobic or neutral residues are tolerated at buried sites in proteins, undoubtedly because of the large favorable contribution of the hydrophobic effect to protein stability (*19*). For example, Fig. 2 shows the results of genetic studies used to investigate the substitutions allowed at residue positions that form the hydrophobic core of the NH_2-terminal domain of λ repressor (*20*). The acceptable core sequences are composed almost exclusively of Ala, Cys, Thr, Val, Ile, Leu, Met, and Phe. The acceptability of many different residues at each core position presumably reflects the fact that the hydrophobic effect, unlike hydrogen bonding, does not depend on specific residue pairings. Although it is possible to imagine a hypothetical core structure that is stabilized exclusively by residues forming hydrogen bonds and salt bridges, such a core would probably be difficult to construct because hydrogen bonds require pairing of donors and acceptors in an exact geometry. Thus the repertoire of possible structures that use a polar core would probably be extremely limited (*21*). Polar and charged residues are occasionally found in the cores of proteins, but only at positions where their hydrogen bonding needs can be satisfied (*22*).

The cores of most proteins are quite closely packed (*23*), but some volume changes are acceptable. In λ repressor, the overall core volume of acceptable sequences can vary by about 10%. Changes at individual sites, however, can be considerably larger. For example, as shown in Fig. 2, both Phe and Ala are allowed at the same core position in the appropriate sequence contexts. Large volume changes at individual buried sites have also been observed in phylogenetic studies, where it has been noted that the size decreases and increases at interacting residues are not necessarily related in a simple complementary fashion (*5, 7, 17*). Rather, local volume changes are accommodated by conformational changes in nearby side chains and by a variety of backbone movements.

The Informational Importance of the Core

With occasional exceptions, the core must remain hydrophobic and maintain a reasonable packing density. However, since the core is composed of side chains that can assume only a limited number of conformations (*24*), efficient packing must be maintained without steric clashes. How important are hydrophobicity, volume, and steric complementarity in determining whether a given sequence can form an acceptable core? Each factor is essential in a physical sense, as a stable core is probably unable to tolerate unsatisfied hydrogen bonding groups, large holes, or steric overlaps (*25*). However, in an informational sense, these factors are not equivalent. For example, in experiments in which three core residues of λ repressor were mutated simultaneously, volume was a relatively unimportant informational constraint because three-quarters of all possible combinations of the 20 naturally occurring amino acids had volumes within the range tolerated in the core, and yet most of these sequences were unacceptable (*20*). In contrast, of the sequences that contained only

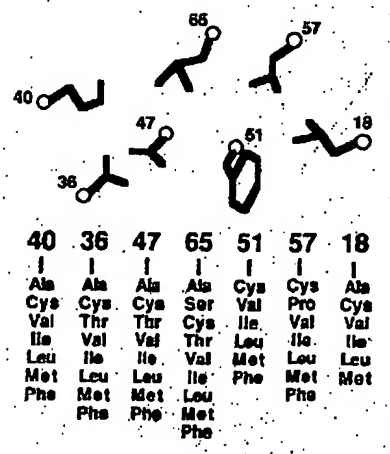

Fig. 2. Amino acid substitutions allowed in the core of λ repressor. The wild-type side chains are shown pictorially in the approximate orientation seen in the crystal structure (*43*). The lists of allowed substitutions at each position are shown below the wild-type side chains. These substitutions were identified by randomly mutating one to four residues at a time by using a cassette method and applying a functional selection (*20*). Not all substitutions are allowed in every sequence background.

the appropriate hydrophobic residues, a significant fraction were acceptable. Hence, the hydrophobicity of a sequence contains more information about its potential acceptability in the core than does the total side chain volume. Steric compatibility was intermediate between volume and hydrophobicity in informational importance.

The Informational Importance of Surface Sites

We have noted that many surface sites can tolerate a wide variety of side chains, including hydrophilic and hydrophobic residues. This result might be taken to indicate that surface positions contain little structural information. However, Bashford *et al.*, in an extensive analysis of globin sequences (*4*), found a strong bias against large hydrophobic residues at many surface positions. At one level, this may reflect constraints imposed by protein solubility, because large patches of hydrophobic surface residues would presumably lead to aggregation. At a more fundamental level, protein folding requires a partitioning between surface and buried positions. Consequently, to achieve a unique native state without significant competition from other conformations, it may be important that some sites have a decided preference for exterior rather than interior positions. As a result, many surface sites can accept hydrophobic residues individually, but the surface as a whole can probably tolerate only a moderate number of hydrophobic side chains.

Identification of Residue Roles from Sets of Sequences

Often, a protein of interest is a member of a family of related sequences. What can we infer from the pattern of allowed substitutions at positions in sets of aligned sequences generated by genetic or phylogenetic methods? Residue positions that can accept a number of different side chains, including charged and highly polar residues, are almost certain to be on the protein surface. Residue positions that remain hydrophobic, whether variable or not, are likely to be buried within the structure. In Fig. 3, those residue positions in λ repressor that can accept hydrophilic side chains are shown in orange and those that cannot accept hydrophilic side chains are shown in green. The obligate hydrophobic positions define the core of the structure, whereas positions that can accept hydrophilic side chains define the surface.

Functionally important residues should be conserved in sets of active sequences, but it is not possible to decide whether a side chain is functionally or structurally important just because it is invariant or conserved. To make this distinction requires an independent assay of protein folding. The ability of a mutant protein to maintain a stably folded structure can often be measured by biophysical techniques, by susceptibility to intracellular proteolysis (*26*), or by binding to antibodies specific for the native structure (*27, 28*). In the latter cases, it is possible to screen proteins in mutated clones for the ability to fold even if these proteins are inactive. Sets of sequences that allow formation of a stable structure can then be compared to the sets that allow both folding and function, with the active site or binding residues being those that are variable in the set of stable proteins but invariant in the set of functional proteins. The DNA-binding residues of Arc repressor were identified by this method (*8*). The receptor-binding residues of human growth hormone were also identified by comparing the stabilities and activities of a set of mutant sequences (*28*). However, in this case, the mutants were generated as hybrid sequences between growth hormone and related hormones with different binding specificities.

Implications for Structure Prediction

At present, the only reliable method for predicting a low-resolution tertiary structure of a new protein is by identifying sequence similarity to a protein whose structure is already known (*29, 30*). However, it is often difficult to align sequences as the level of sequence similarity decreases, and it is sometimes impossible to detect statistically significant sequence similarity between distantly related proteins. Because the number of known sequences is far greater than the number of known structures, it would be advantageous to increase the reach of the available structural information by improving methods for detecting distant sequence relations and for subsequently aligning these sequences based on structural principles. In a normal homology search, the sequence database is scanned with a single test sequence, and every residue must be weighted equally. However, some residues are more important than others and should be weighted accordingly. Moreover, certain regions of the protein are more likely to contain gaps than others. Both kinds of information can be obtained from sequence sets, and several techniques have

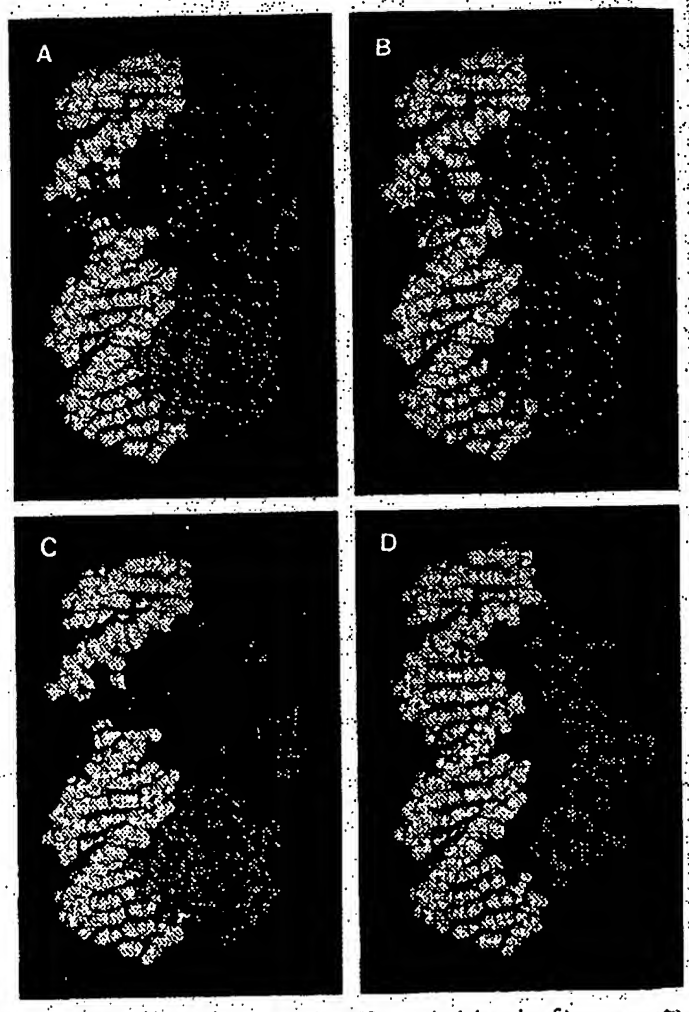

Fig. 3. Tolerance of positions in the NH_2-terminal domain of λ repressor to hydrophilic side chains. The complex (*43*) of the repressor dimer (blue) and operator DNA (white) is shown. In (**A**), positions that can tolerate hydrophilic side chains are shown in orange. The same side chains are shown in (**B**) without the remaining protein atoms. In (**C**), positions that require hydrophobic or neutral side chains are shown in green. These side chains are shown in (**D**) without the remaining protein atoms. About three-fourths of the 92 side chains in the NH_2-terminal domain are included in both (**B**) and (**D**). The remaining positions have not been tested. Data are from (*9, 14, 20, 27, 44*).

been used to combine such information into more appropriately weighted sequence searches and alignments (*31*). These methods were used to align the sequences of retroviral proteases with aspartic proteases, which in turn allowed construction of a three-dimensional model for the protease of human immunodeficiency virus type 1 (*29*). Comparison with the recently determined crystal structure of this protein revealed reasonable agreement in many areas of the predicted structure (*32*).

The structural information at most surface sites is highly degenerate. Except for functionally important residues, exterior positions seem to be important chiefly in maintaining a reasonably polar surface. The information contained in buried residues is also degenerate, the main requirement being that these residues remain hydrophobic. Thus, at its most basic level, the key structural message in an amino acid sequence may reside in its specific pattern of hydrophobic and hydrophilic residues. This is meant in an informational sense. Clearly, the precise structure and stability of a protein depends on a large number of detailed interactions. It is possible, however, that structural prediction at a more primitive level can be accomplished by concentrating on the most basic informational aspects of an amino acid sequence. For example, amphipathic patterns can be extracted from aligned sets of sequences and used, in some cases, to identify secondary structures.

If a region of secondary structure is packed against the hydrophobic core, a pattern of hydrophobic residues reflecting the periodicity of the secondary structure is expected (*33, 34*). These patterns can be obscured in individual sequences by hydrophobic residues on the protein surface. It is rare, however, for a surface position to remain hydrophobic over the course of evolution. Consequently, the amphipathic patterns expected for simple secondary structures can be much clearer in a set of related sequences (*6*). This principle is illustrated in Fig. 4, which shows helical hydrophobic moment plots for the Antennapedia homeodomain sequence (Fig. 4A) and for a composite sequence derived from a set of homologous homeodomain proteins (Fig. 4B) (*35*). The hydrophobic moment is a simple measure of the degree of amphipathic character of a sequence in a given secondary structure (*34*). The amphipathic character of the three α-helical regions in the Antennapedia protein (*36*) is clearly revealed only by the analysis of the combined set of homeodomain sequences. The secondary structure of Arc repressor, a small DNA-binding protein, was recently predicted by a similar method (*8*) and confirmed by nuclear magnetic resonance studies (*37*).

The specific pattern of hydrophobic and hydrophilic residues in an amino acid sequence must limit the number of different structures a given sequence can adopt and may indeed define its overall fold. If this is true, then the arrangement of hydrophobic and hydrophilic residues should be a characteristic feature of a particular fold. Sweet and Eisenberg have shown that the correlation of the pattern of hydrophobicity between two protein sequences is a good criterion for their structural relatedness (*38*). In addition, several studies indicate that patterns of obligatory hydrophobic positions identified from aligned sequences are distinctive features of sequences that adopt the same structure (*4, 29, 38, 39*). Thus, the order of hydrophobic and hydrophilic residues in a sequence may actually be sufficient information to determine the basic folding pattern of a protein sequence.

Although the pattern of sequence hydrophobicity may be a characteristic feature of a particular fold, it is not yet clear how such patterns could be used for prediction of structure de novo. It is important to understand how patterns in sequence space can be related to structures in conformation space. Lau and Dill have approached this problem by studying the properties of simple sequences composed only of H (hydrophobic) and P (polar) groups on two-dimensional lattices (*40*). An example of such a representation is shown in Fig. 5. Residues adjacent in the sequence must occupy adjacent squares on the lattice, and two residues cannot occupy the same space. Free energies of particular conformations are evaluated with a single term, an attraction of H groups. By considering chains of ten residues, an exhaustive conformational search for all 1024 possible sequences of H and P residues was possible. For longer sequences only a representative fraction of the allowed sequence or conformation space could be explored. The significant results were as follows: (i) not all sequences can fold into a "native" structure and only a few sequences form a unique native structure; (ii) the probability that a sequence will adopt a unique native structure increases with chain length; and (iii) the native states are compact, contain a hydrophobic core surrounded by polar residues, and contain significant secondary structure. Although the gap between these two-dimensional simulations and three-dimensional structures is large, the use of simple rules and sequence representations yields results similar to those expected for real proteins. Three-dimensional lattice methods are also beginning to be developed and evaluated (*41*).

Summary

There is more information in a set of related sequences than in a single sequence. A number of practical applications arise from an analysis of the tolerance of residue positions to change. First, such information permits the evaluation of a residue's importance to the function and stability of a protein. This ability to identify the essential elements of a protein sequence may improve our understanding of the determinants of protein folding and stability as well as protein function. Second, patterns of tolerance to amino acid substitutions of varying hydrophilicity can help to identify residues likely to be buried in a protein structure and those likely to occupy

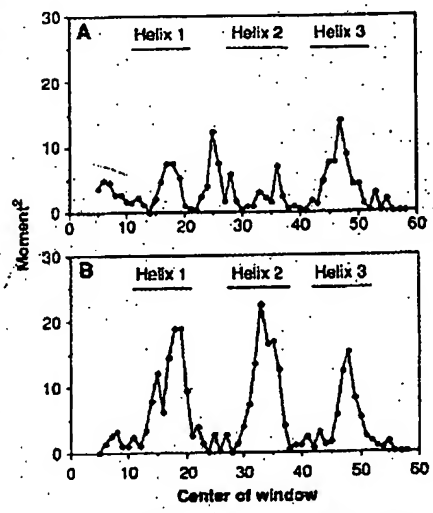

Fig. 4. Helical hydrophobic moments calculated by using (**A**) the Antennapedia homeodomain sequence or (**B**) a set of 39 aligned homeodomain sequences (*35*). The bars indicate the extent of the helical regions identified in nuclear magnetic resonance studies of the Antennapedia homeodomain (*36*). To determine hydrophobic moments, residues were assigned to one of three groups: H1 (high hydrophobicity = Trp, Ile, Phe, Leu, Met, Val, or Cys); H2 (medium hydrophobicity = Tyr, Pro, Ala, Thr, His, Gly, or Ser); and H3 (low hydrophobicity = Gln, Asn, Glu, Asp, Lys, or Arg). For the aligned homeodomain sequences, the residues at each position were sorted by their hydrophobicity by using the scale of Fauchere and Pliska (*45*). Arg and Lys were not counted unless no other residue was found at the position, because they contain long aliphatic side chains and can thereby substitute for nonpolar residues at some buried sites. To account for possible sequence errors and rare exceptions, the most hydrophilic residue allowed at each position was discarded unless it was observed twice. The second most hydrophilic residue was then chosen to represent the hydrophobicity of each position. An eight-residue window was used and the vectors projected radially every 100°. The vector magnitudes were assigned a value of 1, 0, or −1 for positions where the hydrophobicity group was H1, H2, or H3, respectively.

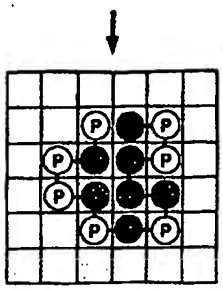

Fig. 5. A representation of one compact conformation for a particular sequence of H and P residues on a two-dimensional square lattice. [Adapted from (*40*), with permission of the American Chemical Society]

surface positions. The amphipathic patterns that emerge can be used to identify probable regions of secondary structure. Third, incorporating a knowledge of allowed substitutions can improve the ability to detect and align distantly related proteins because the essential residues can be given prominence in the alignment scoring.

As more sequences are determined, it becomes increasingly likely that a protein of interest is a member of a family of related sequences. If this is not the case, it is now possible to use genetic methods to generate lists of allowed amino acid substitutions. Consequently, at least in the short term, it may not be necessary to solve the folding problem for individual protein sequences. Instead, information from sequence sets could be used. Perhaps by simplifying sequence space through the identification of key residues, and by simplifying conformation space as in the lattice methods, it will be possible to develop algorithms to generate a limited number of trial structures. These trial structures could then, in turn, be evaluated by further experiments and more sophisticated energy calculations.

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. R. E. Dickerson, *Sci. Am.* **242**, 136 (March 1980).
3. M. D. Hampsey, G. Das, F. Sherman, *FEBS Lett.* **231**, 275 (1988).
4. D. Bashford, C. Chothia, A. M. Lesk, *J. Mol. Biol.* **196**, 199 (1987).
5. A. M. Lesk and C. Chothia, *ibid.* **136**, 225 (1980).
6. M. F. Perutz, J. C. Kendrew, H. C. Watson, *ibid.* **13**, 669 (1965).
7. C. Chothia and A. M. Lesk, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 399 (1965).
8. J. U. Bowie and R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152 (1989).
9. J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241**, 53 (1988); *Proteins Struct. Funct. Genet.*, in press.
10. D. Shortle, *J. Biol. Chem.* **264**, 5315 (1989).
11. J. H. Miller *et al.*, *J. Mol. Biol.* **131**, 191 (1979).
12. S. Sprang *et al.*, *Science* **237**, 905 (1987); C. S. Craik, S. Roczniak, C. Largman, W. J. Rutter, *ibid.*, p. 909.
13. H. C. M. Nelson and R. T. Sauer, *J. Mol. Biol.* **192**, 27 (1986).
14. M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5685 (1984).
15. T. Alber, D. Sun, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
16. D. Shortle and A. K. Meeker, *Proteins Struct. Funct. Genet.* **1**, 81 (1986).
17. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **160**, 325 (1982).
18. W. R. Taylor, *ibid.* **188**, 233 (1986).
19. W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959); R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069 (1986).
20. W. A. Lim and R. T. Sauer, *Nature* **339**, 31 (1989); in preparation.
21. Lesk and Chothia (*5*) have argued that a protein core composed solely of hydrogen-bonded residues would also be inviable on evolutionary grounds, as a mutational change in one core residue would require compensating changes in any interacting residue or residues to maintain a stable structure.
22. T. M. Gray and B. W. Matthews, *J. Mol. Biol.* **175**, 75 (1984); E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **44**, 97 (1984).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
24. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
25. J. T. Kellis, Jr., K. Nyberg, A. R. Fersht, *Biochemistry* **28**, 4914 (1989); W. S. Sandberg and T. C. Terwilliger, *Science* **245**, 54 (1989).
26. A. A. Pakula and R. T. Sauer, *Proteins Struct. Funct. Genet.* **5**, 202 (1989).
27. B. C. Cunningham and J. A. Wells, *Science* **244**, 1081 (1989); R. M. Breyer and R. T. Sauer, *J. Biol. Chem.* **264**, 13848 (1989).
28. B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **243**, 1330 (1989).
29. L. H. Pearl and W. R. Taylor, *Nature* **329**, 351 (1987).
30. W. J. Brown *et al.*, *J. Mol. Biol.* **42**, 65 (1969); J. Greer, *ibid.* **153**, 1027 (1981); J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
31. W. R. Taylor, *Protein Eng.* **2**, 77 (1988).
32. M. A. Navia *et al.*, *Nature* **337**, 615 (1989).
33. M. Schiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967); V. I. Lim, *J. Mol. Biol.* **88**, 857 (1974); *ibid.*, p. 873.
34. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1982); D. Eisenberg, D. Schwarz, M. Komaromy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984); D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
35. T. R. Burglin, *Cell* **53**, 339 (1988).
36. G. Otting *et al.*, *EMBO J.* **7**, 4305 (1988).
37. J. N. Breg, R. Boelens, A. V. E. George, R. Kaptein, *Biochemistry* **28**, 9826 (1989); M. G. Zagorski, J. U. Bowie, A. K. Vershon, R. T. Sauer, D. J. Patel, *ibid.*, p. 9813.
38. R. M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171**, 479 (1983).
39. J. U. Bowie, N. D. Clarke, C. O. Pabo, R. T. Sauer, *Proteins Struct. Funct. Genet.*, in preparation.
40. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
41. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2668 (1989); A. Kolinski, J. Skolnick, R. Yaris, *Biopolymers* **26**, 937 (1987); D. G. Covell and R. L. Jernigan, *Biochemistry*, in press.
42. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
43. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
44. R. M. Breyer, thesis, Massachusetts Institute of Technology, Cambridge (1988).
45. J.-L. Fauchère and V. Pliska, *Eur. J. Med. Chem.-Chim. Ther.* **18**, 369 (1983).

46. We thank C. O. Pabo and S. Jordan for coordinates of the NH_2-terminal domain of λ repressor and its operator complex. We also thank P. Schimmel for the use of his graphics system and J. Burnbaum and C. Francklyn for assistance. Supported in part by NIH grant AI-15706 and predoctoral grants from NSF (J.R.-O.) and Howard Hughes Medical Institute (W.A.L.).

antibodies to YTS 169.4 and YTS 191.1, G. Butcher for antibody to P 7/7, P. Crocker for antibodies to M1/70 and SER-4, and M. Shepard for the murine rIFN-γ. Supported in part by Naval Medical Research and Development Command, work unit 3M162770A870AN121 (S.L.H., G.L., A.S., and M.S.), by U.S. Agency for International Development contract DTE0453C003051-00 (M.R.H. and L.W.), and by Office of Naval Research contract N00014-83-C-0355.

23 December 1988; accepted 7 April 1989

High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis

BRIAN C. CUNNINGHAM AND JAMES A. WELLS*

A strategy, called alanine-scanning mutagenesis, was used to identify specific side chains in human growth hormone (hGH) that strongly modulate binding to the hGH receptor cloned from human liver. Single alanine mutations (62 in total) were introduced at every residue contained within the three discontinuous segments of hGH (residues 2 to 19, 54 to 74, and 167 to 191) that have been implicated in receptor recognition. The alanine scan revealed a cluster of a dozen large side chains that when mutated to alanine each showed more than a four times lower binding affinity to the hGH receptor. Many of these residues that promote binding to the hGH receptor are altered in homologs of hGH (such as placental lactogens and prolactins) that do not bind tightly to the hGH receptor. The overall folding of these mutant proteins was indistinguishable from that of the wild-type hGH, as determined by strong cross-reactivities with seven different conformationally sensitive monoclonal antibodies. The alanine scan also identified at least one side chain, Glu^{174}, that hindered binding because when it was mutated to alanine the receptor affinity increased by more than a factor of four.

SYSTEMATIC REPLACEMENT OF SEGments (7 to 30 residues in length) of human growth hormone (hGH) with sequences derived from nonbinding growth hormone homologs (homolog-scanning mutagenesis) defined a binding patch on a structural model of hGH that included the NH_2-terminal portion of helix 1, a loop between residues 54 and 74 and the COOH-terminal portion of helix 4 (*1*). This analysis provided a general outline of the receptor binding site, but did not identify the specific residues involved in receptor binding. Here, side chains that are important for modulating binding are located by sequential replacement of residues encompassed in the binding patch with alanine. Alanine was chosen as the replacement residue because it eliminates the side chain beyond the β carbon yet does not alter the main-chain conformation (as can glycine or proline) nor does it impose extreme electrostatic or steric effects. Furthermore, alanine is the most abundant amino acid and is found frequently in both buried and exposed positions and all variety of secondary structures (*2*). Alanine-scanning mutagenesis generates a small and systematic set of mutant proteins that can be readily assayed by quantitative binding analysis and avoids the necessity of sorting a library of random mutants by a genetic screen or selection.

A total of 62 single alanine mutants were produced (Table 1) by restriction-selection (*3*) to efficiently enrich for the mutant sequence after primer-directed mutagenesis on a synthetic hGH gene template (*1*). The mutant hormones were expressed in a secreted form from *Escherichia coli* (*4*), and their binding constants were determined for the extracellular portion of the cloned hGH liver receptor by Scatchard analysis (*1*, *5*). This receptor fragment is highly soluble and retains high affinity and specificity for hGH (*5*). The use of the purified truncated receptor in binding assays avoided artifacts associated with binding to receptors on membranes or whole cells.

Overall, the results of the alanine scan (Fig. 1) are consistent with those from the homolog scan (*1*) in showing that the middle and COOH-terminal segments are more important in binding than the NH_2-terminal segment. The largest reductions in binding (~20 times lower) occurred for specific alanine substitutions within the 54 to 74 loop and the COOH-terminal sequence 167 to 191. We extended the alanine scan to include residues 2 to 19 because of uncertainties in the positions of the NH_2-terminal residues in the porcine GH (pGH) structure (*6*). In this segment, alanine substitutions caused more modest reductions in binding; the largest reduction (~6 times lower) was for F10A. For one mutant (E174A) that is located near a number of disruptive alanine mutations (Fig. 1), the affinity for the receptor was actually increased (4.5 times higher).

The most disruptive alanine substitutions form a patch that extends from F10 to R64 and from D171 to V185 (Fig. 2). These side chains appear to be facing in the same direction in the model of hGH. For example, all of the alanine mutants tested in helix 4 that most affect binding (D171A, K172A, E174A, F176A, I179A, C182A, and R183A) are confined to three and one-half turns of this helix, and their side chains

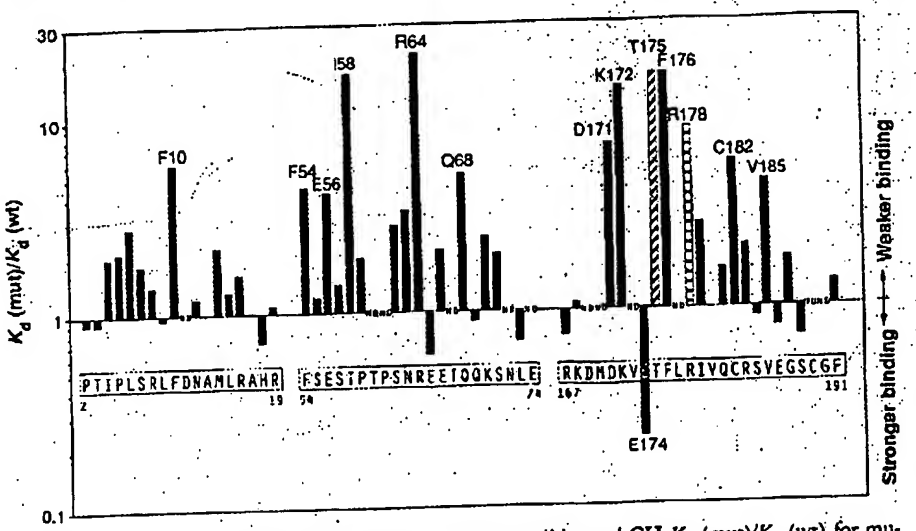

Fig. 1. The change in the dissociation constant relative to wild-type hGH K_d (mut)/K_d (wt) for mutating residues to alanine (black bars), serine (crosshatched bar), or asparagine (hatched bar) within the three binding site segments (Table 1). ND indicates that the K_d (mut) was not determined because of poor expression of the corresponding alanine mutant protein. Residues in which mutations cause a fourfold or greater change in the dissociation constant are labeled. The values are all at or below ±25% (SD).

Department of Biomolecular Chemistry, Genentech, 460 Point San Bruno Boulevard, South San Francisco, CA 94080.

*To whom correspondence should be addressed.

project from the same face of the helix (Fig. 2). From this model, it would appear that T175 is involved in binding because it occupies a central position. Although the T175A mutant could not be expressed in high enough yields from cultures grown in shake flasks to be assayed, a more conservative mutant (T175S) could be. Indeed, the T175S mutant caused a 16-fold reduction in affinity for the hGH receptor (Fig. 1). Similarly, although R178A was poorly expressed, a more conservative mutant (R178N) could be expressed and exhibited an eightfold reduction in receptor binding affinity. The next most disruptive mutant in the COOH-terminus was V185A. Although V185A is outside of helix 4, it is predicted by the model to face in the same direction as the disruptive mutations within helix 4. By comparison, alanine mutations tested outside the binding patch (R167A, K168A, E186A, S188A, and F191A), or within it but facing in the opposite direction from those above (V180A, Q181A, and S184A) generally had little or no effect on receptor binding.

A similar interpreration can be made for alanine mutants in helix 1, although the reductions in binding affinity are more moderate. Within the helix, the alanine substitutions that most disrupt binding are at residues 6, 10, and 14, which are located on the same face of the helix. The least disruptive alanine mutations (L9A, N12A, and L15A) are located on the opposite face of helix 1. This is further confirmed by the fact that monoclonal antibodies (Mab's) 3 and 4, which do not compete with the receptor for binding to hGH (*1*), are strongly disrupted by the N12A mutation (Table 2).

The relative positions of side chains within the 54 to 74 loop cannot be fixed from the model as they can be for those within helices 1 and 4. However, there is a striking periodicity in the binding data in which mutations of even numbered residues cause large reductions in binding relative to odd numbered residues. This is especially true for the first part of this region (54 to 59) and may reflect a structure in which even numbered residues project toward the receptor and odd ones away.

Although it is possible that some or all of the alanine mutations that disrupt the receptor binding do so by causing the molecule to be globally misfolded, it is unlikely for the following reasons. (i) Eight different Mab's to hGH (anti-hGH) whose epitopes are distinct from each other (*1*) and seven of which (Mab's 2 to 8) are sensitive to the folded structure of hGH (*7, 8*) bind as tightly with almost all of the alanine mutants that disrupt binding to the receptor as they do with hGH (Table 2). The exceptions are R64A and C182A, which selectively disrupt binding to the Mab's 6 and 5, respectively. It is likely that these mutations disrupt binding determinants in common between the receptor and Mab's 5 and 6 because these Mab's compete with the receptor for binding to hGH (*1*). As additional controls, two alanine mutants are shown that do not affect receptor binding; one of these (N12A) affects the binding of two Mab's and the other (K168A) affects none of the Mab's. Together these data suggest that binding to either the Mab's or receptor is disrupted by a very local perturbation in the mutant structure, the most obvious being the side chain replacement. (ii) The far ultraviolet circular dichroic spectra for seven of these mutants that have been analyzed so far are virtually identical to that of wild-type hGH (*9*). These observations are consistent with crystallographic data showing that single amino acid substitutions in proteins generally cause only small and local structural perturbations apart from the altered side chain (*10*). We cannot exclude the possibility that small perturbations in structure, which are propagated beyond the side chain substitution (*11*), may affect receptor binding (*12*). Thus, the alanine scan identifies side chains in hGH that can most affect binding; such residues are likely but not necessarily in direct contact with the receptor.

About 20% of the alanine mutants (D11A, T60A, P61A, T67A, N72A, E74A, D169A, M170A, V173A, T175A, L177A, R178A, C189A, and G190A) were not secreted at high enough concentrations in shake flask cultures to be isolated and analyzed (Table 1). Each mutant gene was expressed in the same vector and expression was independent of the specific alanine co-

Table 1. Dissociation constants measured for alanine substitutions of residues 2 to 19 in hGH. Site-specific mutagenesis was carried out on a single-stranded template (pB0475) that had unique restriction sites distributed uniformly about every 15 codons throughout a synthetic hGH gene (*1*). Synthetic oligonucleotides (generally 20 to 40 bases long) that coded for the desired alanine (or other) substitution and altered the closest singly occurring restriction site, were used to prime heteroduplex synthesis (*3*). Mutations between codons 2 to 12, 13 to 19, 54 to 60, 61 to 69, or 70 to 74 were generated by primers that hybridized over and altered the restriction sites Sal I, Apa I, Cla I, Nru I, or Sac I starting at codons 7, 16, 56, 65, or 74, respectively. For mutations between codons 167 to 191, we first introduced a Kpn I site at codon 179 that was subsequently altered by restriction selection (*3*). Heteroduplexes were transformed into *E. coli* BMH 71-18 *mutL* (*23*) and the mixture was grown in LB broth (Luria broth) plus ampicillin (50 μg/ml). The mutant sequence was enriched from a mini-lysate of DNA (*24*) by restriction with the enzyme in which the restriction site was altered by the oligonucleotide (*3*). Residual undigested DNA was transformed directly into *E. coli* JM101, and the correct mutants were identified by dideoxy sequencing (*25*). The combined mutagenesis efficiency for recovery of the 64 correct mutants (from ~150 total clones analyzed was about 60%). Mutant and wild-type GH were secreted from *E. coli* W3110 grown in 20 ml of minimal media containing low phosphate for 24 hours at 37°C in 250-ml shake flasks; periplasmic extracts were prepared by osmotic shock (*4*). The growth hormones were purified uniformly to about 60% homogeneity after an $(NH_4)_2SO_4$ precipitation (final concentration equal to 45% saturation) and their concentrations were determined to precision of SD ± 5% by densitometric scanning of Coomassie-stained SDS gels (*1*). Control experiments showed that the contaminating proteins from *E. coli* extracts had no effect on binding of wild-type hGH to the hGH receptor. The dissociation constant for the soluble portion of the cloned liver hGH receptor was determined by competitive displacement of ^{125}I-labeled hGH and Scatchard analysis (*5*). The standard deviations for K_d were at or generally below ±25% of the values reported. Mutants are abbreviated by the wild-type residue (single letter amino acid designation) followed by its codon number and the mutant residue (typically alanine). Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. NE (not expressed) indicates the mutant protein was expressed at below 2% of wild-type hGH as determined by SDS–polyacrylamide gel electrophoresis or enzyme-linked immunosorbent assay (ELISA) after partial purification from *E. coli* periplasmic extracts.

Mutant	K_d (n*M*)	Mutant	K_d (n*M*)
wt	0.34	Q69A	0.31
P2A	0.31	K70A	0.82
T3A	0.31	S71A	0.68
I4A	0.68	N72A	NE
P5A	0.71	L73A	0.24
L6A	0.95	E74A	NE
S7A	0.61		
R8A	0.48	R167A	0.26
L9A	0.32	K168A	0.37
F10A	2.0	D169A	NE
D11A	NE	M170A	NE
N12A	0.40	D171A	2.4
A13(wt)		K172A	4.6
M14A	0.75	V173A	NE
L15A	0.44	E174A	0.075
R16A	0.51	T175A	NE
A17(wt)		T175S	5.9
H18A	0.24	F176A	5.4
R19A	0.37	L177A	NE
		R178A	NE
F54A	1.5	R178N	2.9
S55A	0.41	I179A	0.92
E56A	1.4	V180A	0.34
S57A	0.48	Q181A	0.54
I58A	5.6	C182A	1.9
P59A	0.65	R183A	0.71
T60A	NE	S184A	0.31
P61A	NE	V185A	1.5
S62A	0.95	E186A	0.27
N63A	1.12	G187A	0.61
R64A	7.11	S188A	0.24
E65A	0.20	C189A	NE
E66A	0.71	G190A	NE
T67A	NE	F191A	0.47
Q68A	1.8		

don. We therefore believe variations in steady-state expression reflect differences in the amounts of mutant protein secreted or proteolytic degradation (or both) of the secreted mutant protein. Mutations in buried positions are often destabilizing (*13*) and can lead to proteolysis in vivo (*14*). Indeed, several of the nonexpressing alanine mutants in helix 4 are located on its hydrophobic face (M170A, V173A, and L177A) (Fig. 2). It is likely that mutation of these residues contributed to the instability of an hGH-pGH hybrid that contained residues 1 to 163 of hGH followed by 164 to 191 of pGH (*1*). However, this is not a general effect because several alanine substitutions were tolerated on the hydrophobic face of helix 1 (L6A, L9A, and F10A) and helix 4 (F176A and V180A). In addition to mutations in hydrophobic residues, sometimes replacing charged or neutral side chains with alanine impaired expression (D11A, T60A, T67A, N72A, E74A, D169A, T175, and R178A). Disruption of the disulfide loop by C182A or C189A mutations leads to very low expression. Several nonexpressing alanine mutants (T60A, P61A, and T67A) were also located in a loop structure containing residues 60 to 67. Thus, although the phenotype of low or nonexpression is characteristic of the mutant protein, this can result from a multitude of structural effects. A more complete interpretation of these results must await a high-resolution structure of hGH and detailed analysis of the rate of secretion and degradation of these molecules in vivo. Moreover, because we have not tested these poorly expressed alanine mutants we do not know whether these wild-type residues are also important for receptor binding.

In two cases tried (T175S and R178N), more conservative substitutions that preserved the hydrogen bonding functional group could be expressed at concentrations near wild type. In another case (C182A), it was possible to purify small amounts of the poorly expressed protein from a high cell density fermentor. Neither of these approaches are very practical for overcoming the problems with poor expression for analysis of large numbers of mutant proteins. However, these alternatives can be applied to a limited extent.

Human growth hormone is a member of a family of hormones that includes the growth hormones (GH), placental lactogens (PL), and the prolactins (PRL) (*15*). Unlike most of its homologs, hGH is capable of binding to somatogenic and lactogenic receptors from a wide range of species. From the alanine-scanning mutagenesis experiments, it is now possible to identify some of the sequence changes that impair binding of hPL, pGH, and hPRL to the soluble hGH receptor (Fig. 3). For hPL, the only differences from hGH among residues where alanine mutations cause more than a fourfold reduction in binding are E56D and R64M. Indeed, when these two hPL residues were introduced into hGH they caused a 30-fold reduction in binding to the hGH receptor (*1*). The combination of these mutations plus I4V and I179M (Table 1) can account (*16*) for greater than a 100-fold reduction in binding for hPL to the somatogenic receptor (*17*). Larger sequence changes occur throughout the three discontinous segments encompassing the binding site for pGH and hPRL (Fig. 3). Of the 13 major sites that alanine replacements cause disruption in binding, 3 are changed in pGH and 9 are changed in hPRL. The combination of these changes can account for a large portion of the reduction in binding for pGH and hPRL to the hGH receptor.

Analysis of sequence divergence among the GH-PRL family has been largely inconclusive in identifying the somatogenic receptor binding site (*15*). The basis for this becomes more clear with information gained from the mutagenesis experiments about the receptor binding site. First, the overall sequence identity for hPL, pGH, and hPRL compared to hGH (85, 68, and 23%, respectively) is essentially the same as the identity within the three discontinuous binding regions (84, 58, and 30%, respectively) (Fig. 3) or within the set of 13 residues where mutations are most disruptive to binding (85, 77 and 31%, respectively). In general, functional differences among these molecules is modulated by sequence differences spread throughout the three discontinuous regions of the binding site. Thus, natural variation within the binding site is no more or less than it is over the entire molecule. Although sequence divergence is a good predictor of the variant binding site regions in such rapidly evolving molecules as protease inhibitors (*18*), it is a poor predictor for identifying sequences responsible for functional differences among the growth hormone family.

Alanine-scanning mutagenesis is a more thorough probe of the receptor binding site than is homolog-scanning mutagenesis (*1*). The latter approach entails anlaysis of multiply mutated proteins, whereas by single substitutions the alanine scan reveals which of these are most crucial in binding (Fig. 2). Moreover the alanine scan identifies important side chains that were missed by the homolog scan such as F10, Q68, and K172. The homolog scan usually introduces conservative side chain substitutions and so may not truly test energetic importance as the alanine scan does. For example, converting Arg 64 to Lys (as in pGH and hPRL) does not reduce binding to the receptor, yet

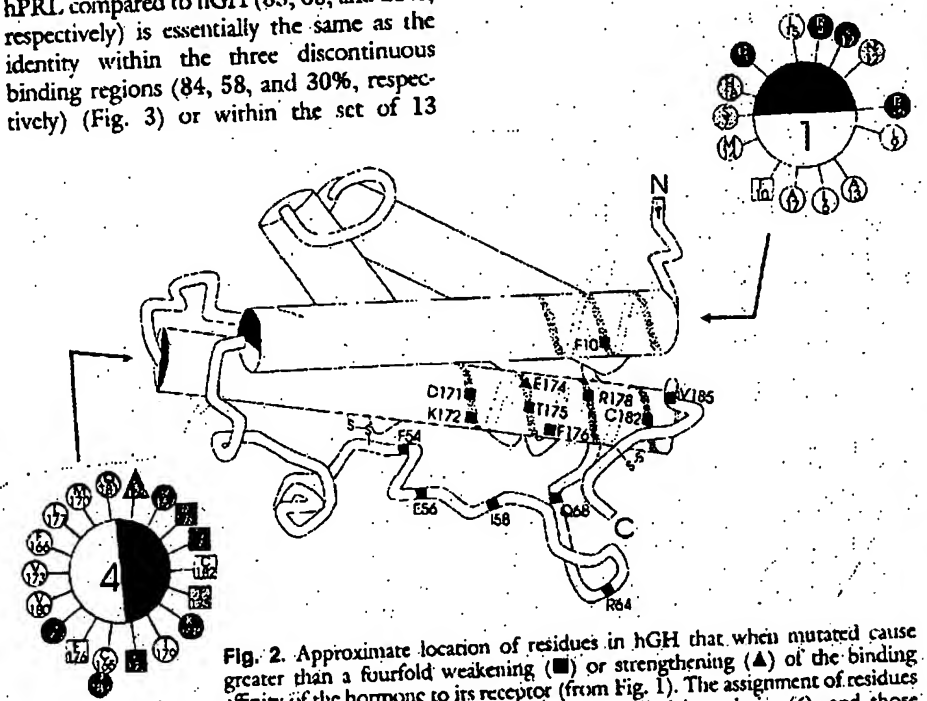

Fig. 2. Approximate location of residues in hGH that when mutated cause greater than a fourfold weakening (■) or strengthening (▲) of the binding affinity of the hormone to its receptor (from Fig. 1). The assignment of residues within each helix was determined by interpolating between the helical boundaries (*6*), and those residues within the loops from the positions of the kinks in the pGH model relative to the disulfides (C53 to C165 and C182 to C189). Helical wheels for helices 1 and 4 are viewed from their corresponding NH_2-terminus (*26*) and include only those helical residues that were mutated (6 to 19 and 165 to 183). These helices are strongly amphipathic (*1*), as shown by the distribution of charged (filled symbols), neutral (shaded symbols), and hydrophobic residues (open symbols). The most hydrophobic side of each helix (nonshaded) is oriented on the model to be facing toward the interior of the protein, which fixes the position of helical residues that most affect receptor binding to be facing outward.

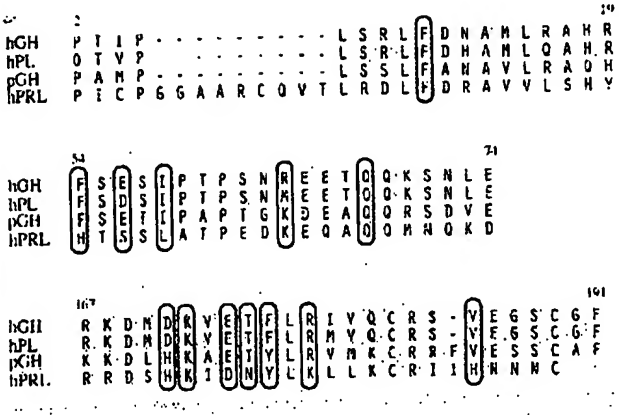

Fig. 3. Comparison of sequences for various members of the GH-PRL family (*15*) within the three segments that strongly modulate binding to the hGH receptor. The shaded residues are identical to hGH and the circled ones are residues that when mutated to alanine (or others) cause the binding constant to change by more than fourfold.

Table 2. Binding of seven different Mab's to hGH and various alanine mutants. An ELISA format was used to determine the concentration (n*M*) of a given Mab necessary to reach 50% saturation in binding to wild-type hGH or various mutants that had been immobilized previously in a microtiter plate (*1*). The standard deviations in these measurements are generally at or below ±30% of the reported value.

Hormone	Mab (n*M*)							
	1	2	3	4	5	6	7	8
hGH	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
F10A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
N12A	0.4	0.4	>75	>50	0.2	0.2	0.08	0.1
I58A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
R64A	0.4	0.4	0.1	0.05	0.2	1.6	0.08	0.1
Q68A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
K168A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
D171A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
K172A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
E174A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
F176A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
C182A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1
V185A	0.4	0.4	0.1	0.05	0.2	0.2	0.08	0.1

conversion of Arg^{64} to Met (as in hPL) reduces binding to the extent of the R64A mutation (*8*). This probably accounts for why the reductions in binding affinity produced by homolog scanning were generally much less than the product of the reductions in binding from component alanine substitutions at the same sites (*16*). For instance, in the mutant pGH (57 to 73) there are ten substitutions but only a 17-fold reduction in binding (*1*). The product of reductions for alanine substitutions at eight of these sites is 273-fold (two of the alanine substitutions were not analyzed because of low expression); most of this discrepancy (~20 times) can be accounted for by the lack of reduction in binding from the R64K substitution as discussed above.

We view the homolog-scanning and alanine-scanning strategies as complementary approaches. The former provides a general location for receptor or antibody epitopes, whereas the latter gives information about specific side chains within a given epitope that are most important for molecular recognition. Both approaches produce a manageable set of mutant molecules that can be analyzed quantitatively. In some instances the mutants could not be expressed in good yield. This problem could sometimes be overcome by selecting more conservative substitutions, or culturing in high cell density fermentors (*1*). In addition, when alanine and glycine side chains are present in the wild-type protein they cannot be probed by mutagenesis except with larger or more conformationally disruptive substitutions, nor can main chain contacts be probed except by potentially disruptive proline substitutions. However, structural analysis of protein-protein interactions have shown that antigenic determinants (*19*) or subunit-subunit interfaces (*20*) are dominated by side chain contacts. In fact, we find the alanine substitutions that most alter binding affinity are replacements of large side chains (both charged and hydrophobic). Finally, disruption of binding by these alanine mutations does not prove a residue is making contact, further structural analysis of the complex is necessary for that.

Although it is anticipated that all epitopes in native globular proteins are to some extent discontinuous (*21*), they invariably contain multiple contacts within a short continuous stretch (<15 residues) of the polypeptide chain (*19*). Thus, once a putative contact region is identified by mutagenesis, chemical modification or another method, the importance of neighboring residues to binding affinity can be assessed by alanine scanning. For example, recently it was reported that alkylation of K172 in hGH substantially reduces binding (*22*). Our data show that other nearby residues in helix 4 are indeed crucial in receptor binding. Finally, alanine scanning has identified at least one side chain in hGH (E174) that naturally disrupts binding to the hGH receptor. Thus, it is likely that the binding between hGH and its somatogenic receptor may be strengthened by optimizing this and other residues that modulate binding. Alanine-scanning mutagenesis is a systematic functional analysis that should be of general use to identify (and subsequently engineer) the side chains that most strongly modulate the interaction between a polypeptide hormone and its receptor (or in other protein-ligand complexes).

REFERENCES AND NOTES

1. B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **243**, 1330 (1989).
2. C. Chothia, *J. Mol. Biol.* **105**, 1 (1976); G. D. Rose, A. R. Geselowitz, G. L. Lesser, R. H. Lee, M. H. Zehfus, *Science* **229**, 834 (1985); M. H. Klapper, *Biochem. Biophys. Res. Commun.* **78**, 1018 (1977).
3. J. A. Wells, B. C. Cunningham, T. P. Graycar, D. A. Estell, *Philos. Trans. R. Soc. London A* **317**, 415 (1986).
4. C. N. Chang, M. Rey, B. Bochner, M. Heyneker, G. Gray, *Gene* **55**, 189 (1987).
5. D. W. Leung *et al.*, *Nature* **330**, 537 (1987); S. A. Spencer *et al.*, *J. Biol. Chem.* **263**, 7862 (1988); G. Fuh, D. Light, J. Wells, unpublished results.
6. S. S. Abdel-Meguid *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6434 (1987).
7. Mutagenic analysis (*1*) has shown that the epitopes for Mab's 2 through 6 are highly discontinuous whereas Mab's 1, 7, and 8 contain more contiguous sequence determinants. Furthermore, the binding affinities to a denatured peptide fragment of hGH (residues 1 to 141) for Mab's 2 through 8 are reduced by more than 40-fold; Mab 1 binds to this fragment with somewhat higher affinity than it does to native hGH (*8*). The 1 to 141 fragment contains all the known sequence determinants for Mab's 1 through 4, 7, and 8 (*1*).
8. B. C. Cunningham and J. A. Wells, unpublished data.
9. M. Mulkerrin, unpublished results.
10. E. E. Howell, J. E. Villafranca, M. S. Warren, S. J. Oatley, J. Kraut, *Science* **231**, 1123 (1986); B. A. Katz and A. Kossiakoff, *J. Biol. Chem.* **261**, 15480 (1986); P. N. Bryan *et al.*, *Prot. Struct. Funct. Genet.* **1**, 326 (1986); T. Alber *et al.*, *Nature* **330**, 41 (1987); S. Sprang *et al.*, *Science* **237**, 905 (1987); R. R. Bott *et al.*, *J. Biol. Chem.* **263**, 7895 (1988). It is possible that larger structural perturbations in mutant proteins have not been observed because such proteins are inherently more difficult to obtain (due to poor expression) or because these proteins are more difficult to crystallize in the same space group.
11. J. A. Wilde *et al.*, *Biochemistry* **27**, 4127 (1988).
12. For example, we hesitate to attribute the effect of the C182A mutant directly on binding (even though its conformation is not drastically altered) because of complications in interpreting the breakage of a disulfide bond.
13. T. Alber, S. Dao-pin, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987); J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241**, 53 (1988).
14. A. A. Pakula, V. B. Young, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8829 (1986); D. Shortle and

B. Lin, *Genetics* 110, 539 (1985).
15. C. S. Nicoll *et al.*, *Endocr. Rev.* 7, 169 (1986).
16. For single mutations at independent and noninteractive sites (x and y), the product of the changes in dissociation constants relative to wild-type [K_d (mut)x/K_d (wt)] [K_d (mut)y/K_d (wt)] should closely approximate the change in the dissociation constant for the double mutant at x and y K_d (mut)xy/K_d (wt) [P. J. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, *Cell* 38, 835 (1984)].
17. S. Burstein, M. M. Grumbach, S. L. Kaplan, C. H. Li, *Proc. Natl. Acad. Sci. U.S.A.* 75, 5391 (1978).
18. M. Laskowski *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 52, 545 (1987).
19. A. G. Amit, R. A. Mariuzza, S. E. V. Phillips, R. J. Poljak, *Science* 233, 747 (1986); S. Sheriff *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 84, 8075 (1987); P. M. Colman *et al.*, *Nature* 326, 358 (1987).
20. P. Argos, *Protein Eng.* 2, 101 (1988); J. Janin, S. Miller, C. Chothia, *J. Mol. Biol.* 204, 155 (1988).
21. D. J. Barlow, M. S. Edwards, J. M. Thornton, *Nature* 322, 747 (1987).
22. L. C. Teh and G. E. Chapman, *Biochem. Biophys. Res. Commun.* 150, 391 (1988).
23. B. Kramer, W. Kramer, H.-J. Fritz, *Cell* 38, 879 (1984).
24. H. C. Birnboim and J. Doly, *Nucleic Acids Res.* 7, 1513 (1979).
25. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463 (1977).
26. M. Schiffer and A. B. Edmundson, *Biophys. J.* 7, 121 (1967).
27. We thank the organic chemistry department at Genentech for providing synthetic oligonucleotides; G. Fuh for making available the purified extracellular portion of the cloned hGH liver receptor; C. Nelson for monoclonal antibodies; P. Carter, L. Abrahmsen, B. Kelley, W. Wood, and L. Presta for critical reading of the manuscript; and W. Anstine for preparing graphics and the manuscript.

29 December 1988; accepted 6 April 1989

Physical Mapping of a Translocation Breakpoint in Neurofibromatosis

JANE W. FOUNTAIN, MARGARET R. WALLACE, MELISSA A. BRUCE, BERND R. SEIZINGER, ANIL G. MENON, JAMES F. GUSELLA, VIRGINIA V. MICHELS, MICHAEL A. SCHMIDT, GORDON W. DEWALD, FRANCIS S. COLLINS*

The gene for von Recklinghausen neurofibromatosis (NF1), one of the most common autosomal-dominant disorders of humans, was recently mapped to chromosome 17 by linkage analysis. The identification of two NF1 patients with balanced translocations that involved chromosome 17q11.2 suggests that the disease can arise by gross rearrangement of the *NF1* locus, and that the NF1 gene might be identified by cloning the region around these translocation breakpoints. To further define the region of these translocations, a series of chromosome 17 Not I–linking clones has been mapped to proximal 17q and studied by pulsed-field gel electrophoresis. One clone, 17L1 (D17S133), clearly identifies the breakpoint in an NF1 patient with a t(1;17) translocation. A 2.3-megabase pulsed-field map of this region was constructed and indicates that the NF1 breakpoint is only 10 to 240 kilobases away from 17L1. This finding prepares the way for the cloning of *NF1*.

VON RECKLINGHAUSEN NEUROFIbromatosis (NF1) is an autosomal-dominant human genetic disease, characterized by café-au-lait spots, multiple neurofibromas that increase in size and number with age, hamartomas of the iris (Lisch nodules), learning disabilities, bone abnormalities, and an increased risk of malignancy (especially glioma and neurofibrosarcoma) (*1*). The incidence of the disease is about 1 in 4000. The specific manifestations and severity are remarkably variable, even within the same family, and the spontaneous mutation rate is high, with 30 to 50% of cases representing new mutations (*2*).

The NF1 gene (*NF1*) has been recently mapped to chromosome 17 by linkage analysis (*3*), and genetic analysis of 142 families by an International Consortium has indicated that *NF1* lies on proximal 17q (*4*). Markers on both sides of the gene have been identified that are within 5 centimorgans of *NF1* (*5*).

In support of this localization, two unrelated patients with NF1 and apparently balanced translocations involving chromosome 17 [t(1;17) and t(17;22)] have been identified (*6*, *7*). In each instance, the chromosome 17 breakpoint is in band q11.2, precisely where *NF1* maps by linkage analysis. Somatic cell hybrids have been constructed that contain the translocation chromosomes from these patients (*7*, *8*). Flanking genetic markers for *NF1* map on opposite sides of the translocation breakpoints (*7–10*), supporting the hypothesis that these translocations directly disrupt the gene. Other somatic cell hybrids created by microcell-mediated gene transfer have been used to further define the location of markers around the *NF1* locus (*9*, *10*). The combination of linkage and physical mapping has now ruled out several chromosome 17 genes, including *ERBA1*, *ERBB2*, and *NGFR*, as candidates for direct involvement in NF1 (*9–11*).

Thus *NF1* is an appropriate target for cloning by reverse genetics (*12*). We have reported (*9*) initial results of physical mapping of this region by pulsed-field gel electrophoresis (PFGE). Although these results allowed us to physically connect some of the closely linked genetic markers, gaps were present in the map and none of these markers detected either NF1 translocation breakpoint. Therefore additional markers were required to visualize this region in more detail. The use of linking clones, which are genomic fragments containing rare restriction sites (*13*), has advantages in such an effort: such clones allow convenient construction of a physical map, and often mark the site of expressed genes. We generated a phage library of Not I–linking clones from flow-sorted chromosome 17 material (*14*) and localized these clones using a somatic cell hybrid panel (*9*). The clones that mapped to 17q11 were tested on PFGE blots (*15*) to see whether abnormal fragments were present in DNA from either NF1 patient with a translocation.

Of 16 linking clones studied, one (called 17L1) identified novel PFGE bands in DNA from the t(1;17) NF1 patient. Specifically, a 0.8-kb Not I–Xho I fragment of 17L1, denoted 17L1A, detected abnormal fragments with the enzymes Bss HII, Sac II, and Not I (Fig. 1A) and also with Mlu I. No abnormality was seen with Sfi I, Eag I, or Xho I (*16*).

A potential pitfall in this analysis can arise from variability in cutting of rare restriction sites. This can occur as a result of actual sequence polymorphism (*17*) or, more commonly, as a result of DNA methylation differences (*15*). Strong evidence that such effects cannot account for the data shown here include: (i) No novel PFGE fragments have been seen with 17L1A in Not I analysis of more than 60 normal chromosomes in several different tissues. (ii) These fragments cannot be accounted for by the effects of incomplete digestion; for example, the intentional partial Not I digest in Fig. 1A demonstrates that the t(1;17)-specific band at 550 kb (lane 9) is distinct from the 460- and 680-kb partial fragments in normal individuals (lanes 10 and 11). (iii) The t(1;17) DNA, which also includes a normal chromosome 17, always shows a normal band as well as the abnormal band, in approximately a 1:1 ratio. (iv) The abnor-

J. W. Fountain, M. R. Wallace, M. A. Bruce, F. S. Collins, Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109.
B. R. Seizinger, A. G. Menon, J. F. Gusella, Neurogenetics Laboratory, Massachusetts General Hospital, Boston, MA 02114.
V. V. Michels, M. A. Schmidt, G. W. Dewald, Mayo Clinic, Rochester, MN 55905.

* To whom correspondence should be addressed.